

Europe needs to pull its weight in research

A broader vision of a pan-European research enterprise will remain just that until the continent's nation states become more imaginative in their approaches to collaboration.

After years of discussion, the fragmented nature of Europe's research base continues to jar against the lofty rhetoric about a united Europe. The European Science Foundation, an association of Europe's national research agencies founded in 1974, remains a bit-player on the continent's research scene. And the European Research Area (ERA), conceived in 1999, still means more in diplomatic circles than in scientific ones (see News Feature, page 768).

Yet the total research investments of the United States and Japan continue to far outstrip those of any European nation, and the need for Europe to pool its limited resources has never been greater. It is time for European scientists to take a hard look at the existing mechanisms for collaboration, and to ask if they are delivering.

Supporters of the ERA, which is intended to unite European science, say it is already approaching reality. The European Commission's Framework programmes for research, large research projects such as CERN, the particle physics laboratory near Geneva, innumerable bi- and multilateral projects, and platforms such as Euroforum, a group of leading European research organizations, have, they argue, already transformed the continent's scientific landscape.

Sceptics take issue with this rosy picture. They point out that 95% of science in Europe is still funded and conducted in single nation states. Governments and funding agencies jealously supervise their spheres of influence and the interests of their specific clientèle. And the structure of research and of scientists' careers in, say, Spain, Slovenia, Finland and Britain are utterly distinct from one another.

Furthermore, the incentives for mobility and scientific exchange within Europe remain weak. There are, for instance, more active French scientists in the Boston area than in the entire European

Union (EU) outside France. Meanwhile, in terms of research spending, patents and high-technology exports per capita, the United States and Japan each continue to lead the EU.

A more productive research and education system is therefore needed, and it is in this light that EU leaders have backed the concept of the ERA. But it cannot be imposed by decree; it needs the combined political will and creativity of all the EU member states.

If Europe is serious about improving its science base, it should start by building on existing strengths. Multinational laboratories, such as CERN and the European Synchrotron Radiation Facility in Grenoble, boast some of the world's best scientists. And there is ample potential for Europe to lead in other areas — climate-change research and cell biology are just two examples. Five or six major new facilities, with a strong capacity to train and integrate scientists from countries spending too little on science, would give European science a valuable boost.

The EU must also decide about a central funding agency for Europe, of the type advocated by the European Science Foundation. There are good reasons for creating such an agency, but its design and goals must be carefully defined. It might find a niche, for example, by concentrating on supporting high-risk projects in basic research that are failing to win support under the current system.

In the United States, DARPA, the Defense Advanced Research Projects Agency, has found such a niche, supporting important basic research in various disciplines. It achieves this by seconding outstanding individuals from the universities as its programme managers, and allowing them more leeway in project selection than other research agencies. To progress, European research needs support mechanisms involving less red tape and a greater willingness to tolerate failure. ■

A Goldin legacy

The US space programme will miss the colour, character and energy that NASA's departing administrator brought to his job.

Whatever else might be said about Dan Goldin, who resigned last week as administrator of the US space agency NASA, he was never dull. In every situation, Goldin displayed an occasionally reckless passion that made him one of the most inspiring leaders the agency has had.

Goldin's legacy will doubtless be closely tied to the aphorism, "better, faster, cheaper", which he applied to the agency's missions. For space scientists, the 'faster' element was particularly welcome. Goldin managed to increase the flight rate of small spacecraft, and so rescued space science from the tedium and inherent risk of flying one giant mission every five or ten years.

The verdict is still out on 'better' and 'cheaper', however. NASA engineers of the old school used to say, "better, faster, cheaper — *pick any two*". After more than nine years under Goldin, their scepticism may yet win out. They can, after all, point to NASA's Mars programme,

which became cheaper and faster but lost two spacecraft in the process.

The International Space Station was, of necessity, Goldin's other preoccupation. His supporters will say that he got the project built in difficult circumstances. Construction, so far, has gone remarkably smoothly. But the budget is again out of control, placing the project back in crisis mode — just where it was when Goldin arrived in 1992.

Even if his vision of a streamlined NASA does fall flat, not all of the blame should fall on Goldin. He was a reformer with little real power to reform. Unlike a corporate executive, he couldn't shape his organization as he wished — not with congressional delegations protecting every NASA field centre. So official Washington cheered him on, but never really got behind his efforts to reform the space agency. That was unfortunate for NASA: men such as Goldin don't arrive in government administration very often, and it will be lucky to see his like again. ■

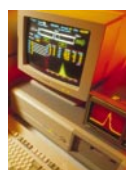




Woolly thinking
Prion-disease study
fails to convince over
risk in sheep
p760



Taking off
NASA's 'faster, better,
cheaper' chief bows
out after nine years
p762



Protein prospects
Researchers gear
up for international
proteomics project
p763



Scales of justice
Zoologists cleared
of faking 'living
fossil' photograph
p765

Storm clouds gather over science as war build-up accelerates

Matthew Davis, Washington

A long period of expansion for civilian science in the United States looks set to end next year, as President George W. Bush's administration shifts its priorities to wage war against terrorism.

Champions of basic research in Washington seem reconciled to the shift. According to congressional sources, it could easily derail existing plans to double the budget of the National Institutes of Health (NIH) between 1998 and 2003, and require real cuts in basic research funding at the National Science Foundation (NSF), the Department of Energy (DOE) and the space agency NASA.

Mitch Daniels, director of the White House Office of Management and Budget, said in a speech last week that for the 2003 budget, which he is now preparing, initiatives not related to the 11 September terrorist attacks have been deemed "lesser priorities".

The Bush administration was already sceptical about investment in non-biomedical basic research before the attacks (see *Nature* **413**, 5; 2001). In a letter to *Nature* (**413**, 566; 2001), however, Daniels stressed that these programmes would be fairly evaluated according to criteria agreed to by scientists.

But since the attacks, many scientists have become increasingly nervous that their research programmes are about to be branded as luxury items by a government that is interested only in essential goods. Their fears are likely to be confirmed when Bush pub-



New agenda: as Capitol Hill sat silent last week, Mitch Daniels (left) was revising budget priorities.



lishes his budget proposal for the 2003 fiscal year, next February.

One lobbyist questions whether supporters of biomedical research can, with the exception of work related directly to bioterrorism, "plausibly or honestly make a case that the NIH and basic biomedical research have a role in fighting terrorism". Several lobbyists and congressional staff say the answer is probably "no" and think there is a strong chance that the effort to double the NIH's budget within five years will stop next year, one year short of completion.

Although the NSF is eager to prove its relevance — last week, its website trumpeted the award of a grant to study "human, social responses" to the attacks — a congressional aide who works on its funding predicts that it

will be judged harshly — and that members of Congress, including those who strongly support basic research, will accept cuts if they are needed to fund a war.

Political analysts say that Bush will be forced to cut expenditure not relating to war in his 2003 budget to avoid a large deficit. Research and development programmes, which account for about one-seventh of the government's 'discretionary' spending, are obvious targets for such cuts.

So the 2002 budgets set to be completed next week may represent the high watermark for science programmes. The NIH seems likely to receive a \$3-billion increase, and the NSF will be boosted by around \$400 million, whereas the DOE and NASA will each get about the same as they did this year. ■

Fallout from terrorism hits summits on developing world

**K. S. Jayaraman, New Delhi,
and Philipp Weis, Munich**

The planned Third World Academy of Sciences (TWAS) meeting and a World Food Summit are among the latest events to be postponed in the aftermath of the 11 September attacks on the United States.

The Indian National Science Academy (INSA) was preparing for the eighth TWAS general conference, set to begin in New

Delhi on 27 October, when it was informed of the decision by the TWAS council.

"TWAS decided to postpone the meeting as most of its members and speakers at the meeting were reluctant to undertake a journey to this part of the world in the prevailing situation," says Govardhan Mehta, INSA's president and chairman of the meeting's organizing committee.

Meanwhile, the food summit, organized by the Food and Agriculture Organization

and expected to attract more than 100 heads of state, has also been postponed, perhaps until next June.

The summit would have been a follow-up to the first World Food Summit, held in 1996, which declared a goal of halving the number of undernourished people in the world by 2015. World leaders had been due to meet in Rome on 5–9 November to "outline the measures they envisage for achieving these goals". ■

Brain mix-up leaves BSE research in turmoil

Declan Butler

In a report that could have plunged the British meat industry back into crisis, parliament was due to receive research results this week confirming that traces of bovine spongiform encephalopathy (BSE) had been found in sheep.

But at the eleventh hour, the Department for Environment, Food and Rural Affairs (DEFRA) declared that the experiments conducted by the Institute for Animal Health (IAH) near Newbury, Berkshire, were suspect.

DEFRA's statement, given on 17 October, revealed the results of tests by the Laboratory of the Government Chemist (LGC) near London. These, the department said, showed that IAH researchers had spent five years and £217,000 (US\$308,000) accidentally testing cow brains instead of sheep brains for BSE. The episode was publicly characterized by John Krebs, chairman of the Food Standards Agency, as a "cock-up".

But since DEFRA's statement was issued, it has emerged that the LGC's findings are themselves open to dispute. Moreover, experts say, the IAH's experiment might never have definitively answered the question of whether BSE is present in sheep.

Indeed, following the government statement, one of its own scientific advisers, who requested anonymity, said that Britain's approach to BSE research was "characterized by a lack of external quality control and peer review". The only thing proven by the episode, another adviser said, was that "there has not been and still isn't any proper, large, well-designed study to ascertain whether BSE exists in sheep".

Masked invader

According to previous experiments done at the IAH (see J. D. Foster, J. Hope & H. Fraser *Vet. Rec.* **133**, 339–341; 1993), sheep fed infected cattle brain can develop BSE. Until a ban in July 1988, British sheep were fed BSE-contaminated meat and bone-meal (see *Nature* **395**, 6–7; 1998). But scrapie — a spongiform encephalopathy similar to BSE, which does not seem to transmit to humans — is endemic in British sheep, and its symptoms are indistinguishable from those of BSE. So it is not known whether BSE has entered the sheep population and is being masked by scrapie.

The experiment at the centre of the controversy began in 1997 as an examination of a slurry of almost 3,000 brains of scrapie-infected sheep, collected by scientists at Britain's now-defunct Ministry of Agriculture, Fisheries and Food.

The tests used a technique, pioneered by Moira Bruce at the IAH's laboratory in Edinburgh, to discriminate between BSE



Unknown quantity: despite research, it remains impossible to say whether British sheep have BSE.

and scrapie. In this method, mice are injected with sheep extracts, and the incubation times and brains of mice that succumb to the disease are analysed. But the whole process can take up to two years.

Some researchers were critical of the experiment design from the outset. Because all of the sheep-brain samples were pooled together, they say, there was the possibility of contamination with cow brains held in the same laboratories, as they were excised on the same slabs. The experiment also lacked controls using samples spiked with BSE.

And although the experiment might have established that feed had infected sheep in the early 1990s — something that would surprise few specialists in the field — it could not have answered the key question of whether enough sheep were infected to trigger contagious spread of BSE in the flock. BSE is not contagious in cattle, but some experts worry that in sheep it might transmit in a similar way to scrapie, and spread through the national flock.

The government had established that the preliminary results of the IAH experiment indicated a prevalence of BSE of between 0.1% and 1% in sheep — even at 0.1% this would have had serious implications for public health. On this basis, according to members of the Spongiform Encephalopathy Advisory Committee (SEAC), which advises the UK government on matters related to such 'prion diseases', the government

was prepared to announce the findings to parliament, until DEFRA revealed that the sheep-brain samples in the experiment were, in fact, cow brains.

But Chris Bostock, director of the IAH, complains of what he terms the "completely unscientific manner" in which the matter has been handled by DEFRA. Despite requests, he says, the institute has yet to obtain copies of the protocols used by the LGC or of its results. Several members of SEAC also say that they have not had an opportunity to review the data.

Sample questions

Although DEFRA's announcement has been portrayed as meaning that the IAH results are flawed, Bostock says that there is merely a "discrepancy" between the LGC's assessment and that of his own scientists. He says that the IAH had earlier carried out its own tests to verify the source of its samples, looking for arginine at position 171 on the prion protein linked to scrapie — a signature considered unique to sheep. The results were positive, he says.

Samples were also sent earlier this year to the government's Central Veterinary Laboratory (CVL) in Weybridge, Surrey, he says, which found no evidence of bovine material. But on 19 October, two days after DEFRA announced its findings, the CVL informed Bostock by e-mail that its earlier results could not be relied upon, as a mix-up of samples had occurred.

In addition, *Nature* has learnt that the samples sent to the LGC frozen in dry ice were delayed and reached the laboratory at room temperature. Whether this affected the outcome of the LGC's tests is uncertain, but one member of SEAC, who had not been informed of this fact, says he remarked on seeing the LGC data that he was surprised that the DNA seemed "highly degraded". Bostock says that frozen duplicates are now "being kept under lock and key", pending the outcome of an internal inquiry into the affair.

Meanwhile, other experiments are under way at the IAH to test 180 individual sheep brains for BSE. But critics say that such a small sample is unlikely to detect BSE at levels of around 0.1%. They argue for diagnostic tests that could test thousands of samples in weeks. From January of next year, the European Union plans to use such techniques to monitor sheep for prion diseases.

Last week's events may yet have regulatory repercussions. On 26 October, the European Union is expected to release new guidelines on public-health controls needed to counter the risk of BSE in sheep. Several advisers to the UK government also argue that existing measures should be reinforced to acknowledge this risk more fully. ■

Grants to help women climb academic ladder

Rex Dalton, San Diego

Complaints that women still face discrimination in the upper reaches of US academia have been mounting in recent years. Now the National Science Foundation (NSF) — whose own director, Rita Colwell, is the first woman ever to hold the position — plans to do something about it.

On 9 October the agency issued eight 'transformation grants', worth over \$3 million each, to help universities try out different approaches to tackle alleged inequities in promotion practice on their campuses.

"There is not an institution in the country that couldn't use this kind of initiative," says Alice Hogan, head of the new NSF programme, which it calls Advance. "This is an investment," she says. "We are concerned about the health and viability of the American scientific and engineering workforce."

During the past 25 years, numerous reports have highlighted problems in hiring women faculty, providing them with tenure, and promoting them to positions such as that of department chair or dean. Many universities have undertaken individual plans to address deficiencies. But interviews with NSF grant winners show that such problems continue to persist.

Particle physicist Tricia Rankin, for example, who will run the Advance programme at

the University of Colorado in Boulder, says: "Twenty years ago, I can remember being told the times are changing. Physics has certainly made improvements, like increasing the number of women physicist graduate students, but it is clear women are not making the breakthroughs anticipated."

The Colorado Advance programme is designed to address campus-wide issues. It seeks "to improve the management and leadership skills of women so they can go up the academic ladder," says Rankin. "But if you really want to encourage change, you are better off modifying the behaviour of men, not women."

Susan Bryant, dean of biological sciences at the University of California, Irvine, who will lead the Advance programme there, says it will concentrate on making sure that women are adequately represented in candidate pools for faculty positions. "People have a tendency to forget either to be inclusive or to make a little extra effort," says Bryant.

At the University of Washington in Seattle, Denice Denton, dean of the school of engineering, will direct the programme. "I would like to see more women department chairs," she says. Women faculty have grown in number at universities such as Denton's, but few women chair departments. The University of Washington has no women chairs in science or engineering, for example.



First lady: will Rita Colwell's success provide a blueprint for future women scientists?

The other Advance programme grants will go to the Georgia Institute of Technology in Atlanta, New Mexico State University at Las Cruces, the University of Puerto Rico, the University of Michigan and the University of Wisconsin at Madison.

The eight winners were selected from 76 universities that applied for the grants, and a ninth winner is expected to be announced shortly.

www.nsf.gov/advance

Proposed scheme will scrutinize student supervisors

David Adam, London

The pay may still be lousy and the hours at the bench long, but the lot of postgraduate students in British universities could be ready to take a turn for the better. Higher-education funding councils are set to introduce new standards and better training for supervisors of research degrees.

Poor supervision is often cited by research students as a problem in British universities. Last year, a review of research policy and funding by the Higher Education Funding Council for England (HEFCE), the government body that distributes teaching and infrastructure funding to English universities, stated that new standards are needed for scientists who act as supervisors. The Quality Assurance Agency for Higher Education, a watchdog body, has also called for better supervisor training.

Some universities and funding agencies already have guidelines in place, but no formal qualifications are required for supervising graduate students. In a move to fill this gap, the Biotechnology and

Biological Sciences Research Council (BBSRC), a life-sciences funding agency, last week announced a new national training and accreditation scheme for them.

The BBSRC scheme has already been piloted at the Institute for Animal Health in Pirbright, Surrey. Supervisors who applied for the qualification were assessed to

determine their ability to plan research projects and to check on student progress. The institute has now made the qualification mandatory for all staff who supervise students.

"Of course there have been bad supervisors in the past but we have to recognize that the entire environment is changing," says Peter Mertens, senior academic tutor at the institute. "It's becoming faster, more competitive and more difficult, and we have to make sure the system is up to that."

The HEFCE, meanwhile, is planning to follow up its report with new standards for supervisors, which it hopes to put in place by 2003. The plans are at an early stage, but John Rushforth, head of research policy at the council, says that universities or departments whose staff fail to measure up will lose money. "If it all works as we envisage, then if you don't meet the standards, you don't get the funding," he says. The BBSRC says that its scheme is voluntary, however, and that there are no plans to link it to funding.



NASA head bows out after a record run

William Triplett, Washington

Dan Goldin has announced his departure from NASA, ending what observers say was a stormy but ultimately successful stint as administrator of the US space agency.

The 61-year-old Bronx-born aerospace engineer revealed on 17 October that he will leave next month. Goldin received plenty of criticism during his record nine-and-a-half-year tenure. But he provided clear vision and immensely energetic leadership at a difficult time for NASA, and the agency's scientists and engineers will now be wondering who will fill his shoes.

Goldin will go down as "the greatest administrator NASA ever had", predicts John Pike, former head of space policy at the Federation of American Scientists and an erstwhile critic of both NASA and its leadership; he now runs the website GlobalSecurity.org.

When Goldin accepted the job of NASA administrator from then-President George Bush in early 1992, he was general manager of TRW, an Ohio-based aerospace and technology contractor. At that time, the agency was still suffering from the aftermath of the 1986 Challenger space-shuttle disaster.

"When Goldin took over, NASA was saddled with problems," says John Logsdon, director of the Space Policy Institute at George Washington University in Washington DC. "He didn't solve all of them, but he did give NASA a strategic vision it badly needed, as well as placing an overall emphasis on cutting-edge technology."

"He was the right man at the right time," says Bradford Parkinson, an aeronautics engineer at Stanford University in California and, until six months ago, chairman of the NASA Advisory Council. "He was an agent of change and NASA needed that because it was a relatively bloated bureaucracy. Since he has a strong sense of himself, he did things others probably couldn't have



Boldly going: departing administrator Dan Goldin brought vision and energy to his role.

done. But he also angered a lot of people."

With little prospect of obtaining increases in NASA's annual budget — which remained stuck at around \$14 billion throughout his tenure — Goldin reduced the agency's direct workforce from some 25,000 to 18,500. He admitted to Congress last year that this reduction was probably too steep.

Meanwhile, the management philosophy he introduced for saving money — "faster, better, cheaper" missions — generated complaints that NASA's work was being dangerously underfunded. Critics

blamed this approach for the loss in 1999 of two probes to Mars.

"A lot of people said 'faster, better, cheaper' was flawed," says Parkinson. "But it wasn't. Rather, it was new thinking, which meant it was bound to have some breakage." Parkinson believes not only that the philosophy is sound, but also that it will prove to be one of Goldin's enduring contributions to the space programme.

Goldin's other memorial may be the survival, albeit in curtailed form, of the International Space Station project. In 1993, the project came within one vote of elimination in the House of Representatives. But Goldin, working closely with vice-president Al Gore, brought Russia in and ensured the project's survival.

The station currently faces fresh cost overruns of \$5 billion, and a report due next month is expected to criticize strongly NASA's management of the project. But Logsdon predicts that the report will concentrate its fire on the Johnson Space Center in Texas, and that Goldin will ultimately be remembered as the man who kept the project alive.

The toughest challenges facing Goldin's successor will include getting space-station costs under control and developing a replacement vehicle for the space shuttle. The aftermath of the 11 September terrorist attacks is likely to intensify pressure on NASA's budget. "The agency was a low priority for this administration before the attacks," says Pike. "It is now a very low priority."

"NASA still needs to sell itself to Congress and to the public," says Parkinson, "but on substance, not stunts. The right leader — say, someone slightly more benevolent — should be able to take the agency to the next step." Logsdon concurs, saying that the agency still needs "a reformer and consolidator" to head it, but that it should be someone "a little calmer" than Goldin. ■

Faculty concerns cost university venture-capital funds

Rex Dalton, San Diego

A proposal for venture-capital funding of research projects at the University of California, San Francisco (UCSF), has been cancelled after an investment agency withdrew its offer of \$10 million to support it.

The UCSF had planned to use the money from the California Public Employees' Retirement System (CalPERS) to award grants of up to \$500,000 to promising research projects to develop drugs or technologies (see *Nature* 413,

95; 2001). Burrill, a venture-capital firm in San Francisco, was going to manage the fund.

But earlier this month, CalPERS officials said they were withdrawing their investment offer. According to Regis Kelly, UCSF's executive vice-chancellor, the withdrawal was linked to concerns raised by UCSF faculty about the programme's potential impact on the university's academic mission. CalPERS officials declined to discuss the decision.

"We are absolutely disappointed," Kelly says. Attempts will be made to address faculty concerns for future ventures, he adds, saying: "We are still interested in setting up a similar fund."

Endocrinologist Daniel Bikle, chair of UCSF's academic senate, who complained that faculty leaders weren't appropriately included in the negotiations with Burrill, says he too is disappointed. But he adds: "We are not a commercial enterprise. Things have to be done without jeopardizing the academic mission." ■

Workshop prepares ground for human proteome project

Alison Abbott

Plans to decipher the human proteome were explored at a meeting earlier this month organized by the recently created Human Proteome Organisation (HUPO).

About 100 researchers and officials gathered in Leesburg, Virginia, on 7 October to discuss a protein equivalent of the Human Genome Project. But participants warned that the initiative could take years to set up.

"The decision to launch the Human Genome Project was made after no less than two years of discussion, followed by three years of pilot effort," said Sam Hanash, a cancer specialist at the University of Michigan, who in June was elected as HUPO's first president. "HUPO knows it is important to proceed step by step."

Understanding the proteome — the pattern of proteins produced by a cell under particular conditions — would accelerate drug discovery, Bernard Schwetz, acting principal deputy commissioner of the Food and Drug Administration, told the workshop.

Participants agreed that HUPO should help to organize an international human proteome project, from which all data would be made freely available — but that there should be a planning and piloting phase before the launch of a full-blown project.

The planning phase will identify model systems — one or two human cell types — and a subset of proteins, such as those representing a specific pathway or subcellular compartment. The three key areas of proteomics — protein expression, protein function and databasing — would then be piloted in these models. "This is a good way to test feasibilities and build up technologies," says Joshua LaBaer, director of the Institute of Proteomics at Harvard Medical School in Boston.

In practice, one pilot project is already under way. This is the Alliance for Cellular Signalling, a multidisciplinary, multi-laboratory initiative that is creating a database of the properties of proteins involved in cellular signalling in certain mouse cells (see *Nature* **402**, 219–220; 1999).

The alliance is keen to cooperate with the proteomics project. "We have a head start on the issue," says Ron Taussig, associate director of the alliance and a cell



No rush: techniques such as mass spectrometry (above) will need to be marshalled for use in proteomics — prompting Sam Hanash (left) to stress the need for patience.

biologist at the University of Texas Southwestern Medical Center in Dallas. "We will make all our databasing knowledge available," he adds.

Unlike the Human Genome Project, which relied on one basic technology for gene sequencing, proteomics uses many new technologies, not all of which have been optimally developed.

Early doubts among some researchers that a single international project could handle the diverse faces of proteomics have been dismissed by participants at the meeting. But Francis Collins, director of the National Human Genome Research Institute in Bethesda, Maryland, cautions that the necessary technologies will have to be identified and proven before the project can be scaled up to high throughput.

Leroy Hood, president of the Institute for Systems Biology in Seattle, Washington, says that for the proteome project to succeed it will have to be both ambitious and tightly focused.

HUPO hopes to assist the project by helping to coordinate activity internationally and by offering advice to potential funders of the work. It also plans to organize training programmes in proteomics technologies. ■

www.hupo.org

Confusion over law leaves stem-cell research in the lurch

Xavier Bosch, Barcelona

A dispute between Spain's ministries of health and science has left the country's stem-cell researchers unsure of the legal status of their work.

About a dozen research groups in Spain are thought to be working with imported human embryonic stem cells, but they face an uncertain future after health-ministry officials suggested in August that one group was in breach of the law. The science ministry has since defended the work, but admits that the law governing the research is unclear and out of date.

The problem arose when Bernat Soria, head of the Bioengineering Institute at Miguel Hernández University in Alicante, spoke about his use of human embryonic stem cells in diabetes research at a conference in July at the University Rey Juan Carlos in Madrid. When the press covered Soria's presentation, they drew attention to the fact that Soria was using human embryonic stem cells. He had obtained the cells from the Israel Institute of Technology in Haifa.

In response to the reports, the health ministry ordered Soria to suspend his work and launched an inquiry. Soria says that health-ministry officials told him that the importation of the cells breached a 1988 law on assisted reproduction. The science ministry defended Soria's research, but has not denied that importing the cells may have been illegal. Soria is now looking for a lawyer to help resolve the issue.

The health ministry's inquiry has since been closed following a statement from Miguel Hernández University that no work using human embryonic stem cells has been carried out on its premises. The statement, and the university's failure to publicly back his work, has angered Soria. "Officially, I have not been conducting such research and I have not used embryonic stem cells," he says.

Ramon Marimon, state secretary at the science ministry, has said that a dozen groups are using stem cells in Spain. The researchers have not been publicly identified, but are all believed to be using imported stem cells, as extracting them may also be illegal under the 1988 law.

A 1998 report for the health ministry, written by the National Commission on Human Assisted Reproduction, recommended that the law be updated to allow extraction "as soon as possible". But divisions in the government are said to have blocked its implementation. ■

Health statistics fuel South Africa's AIDS controversy

Cape Town A report on AIDS by the South African Medical Research Council (MRC) has reignited debate over how the country treats the disease. The long-awaited report, released last week, stated that AIDS is responsible for about a quarter of all deaths in South Africa — making it the biggest single cause of mortality in the country.

President Thabo Mbeki has previously queried government spending plans for AIDS programmes on the basis that AIDS was not the major cause of death. The president was later shown to have used outdated statistics.

But a spokesperson for Statistics SA, the country's central statistical service, criticized the report's methodology even before it was released, and said that Statistics SA was conducting its own study.

Health minister Manto Tshabalala-Msimang also attacked the report, and is said to have warned the MRC researchers that they had "chosen to act in ways which place themselves in a hostile position vis-à-vis the government, and it will be necessary for this serious situation to be attended to".

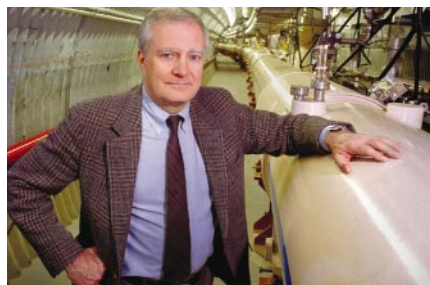
The report was defended by MRC

president Malegapuru Makgoba, who said that its findings were compatible with several other studies, including projections made by Statistics SA.

Committee paves way for Marburger confirmation

Washington The US Senate is set to confirm physicist John Marburger as director of the White House Office of Science and Technology Policy, and science adviser to President George W. Bush.

The Senate Committee on Commerce, Science and Transportation approved Marburger's appointment on 17 October, clearing the way for his confirmation by the full Senate, which could be achieved



Physical effort: Brookhaven's John Marburger prepares to take his skills to the White House.

as early as this week.

Marburger, who has conducted research in lasers and nonlinear optics, directs the Brookhaven National Laboratory on Long Island, New York. A graduate of Princeton and Stanford, he was chair of the department of physics and electrical engineering at the University of Southern California in the 1970s and president of the State University of New York at Stony Brook from 1980 to 1994.

Breast screening 'does not save lives'

Copenhagen A re-analysis of screening trials for breast cancer has concluded that, across populations, there is no evidence that the mammography procedure saves lives.

X-ray imaging is used in Europe and the United States in an effort to catch breast tumours early. Doubts over its effectiveness were raised in January this year by Ole Olsen and Peter Gøtzsche of the Nordic Cochrane Centre in Copenhagen, whose controversial analysis of large-scale screening trials suggested that they did not reduce mortality.

With the help of the respected Cochrane Collaboration, an independent body that supports systematic reviews of healthcare treatments, the pair have now confirmed and extended their conclusions (P. C. Gøtzsche &

O. Olsen, *Lancet* **358**, 1340–1342; 2001).

Among other problems, Olsen and Gøtzsche found that several trials were flawed, and that the benefits of screening may be outweighed by the risk to patients from treatments such as radiotherapy.

► <http://www.cochrane.org>

Japan steps up tests as girl shows vCJD symptoms

Tokyo The Japanese government has launched a massive bovine spongiform encephalopathy (BSE) testing programme in a bid to persuade the public that the country's beef is safe to eat.

Japan's first case of BSE was confirmed last month (see *Nature* **413**, 240; 2001). The government now intends to test all slaughtered cattle for BSE. Samples from around 1.3 million animals are expected to be analysed.

But consumer confidence has been hit further by reports that a teenage girl admitted to a Tokyo hospital is showing symptoms of variant Creutzfeldt–Jakob disease (vCJD), the human brain-wasting disease associated with eating BSE-infected beef products.

The hospital is now testing the girl for vCJD and says that the results will be ready in three months.

NSF puts \$50 million into teaching scheme

San Diego The US National Science Foundation (NSF) is ploughing \$50 million into a scheme to improve the teaching of science and mathematics in schools, from kindergarten age onwards.

Five consortia of universities will receive \$10 million each over five years. The funds will be used to establish learning and technology centres to study science and maths teaching, and to encourage science students to become teachers. "In grades 7–12 [age 11–16 years], approximately 33% of mathematics teachers and 20% of science teachers have neither a major nor a minor in their teaching field," says Judith Ramaley, head of the NSF's education and human resources directorate.

The NSF has already issued similar five-year awards for two prototype centres, and plans to fund three more in the near future.

► <http://nsf.gov/home/ehr>

Coelacanth researchers 'lacked rigour' over photo

Paris The French Institute of Research for Development (IRD) has cleared two of its researchers of falsifying a photo of a coelacanth, a 'living fossil' fish thought to



Fish paste: the coelacanth, photographed while swimming, lies flat in the fake photo (front right).

be extinct until a live specimen was found off the South African coast during the 1930s.

The photo accompanied information sent to *Nature* in which Bernard Séret, Laurent Pouyaud and Georges Serre claimed they had discovered a coelacanth off the Indonesian coast in 1995, three years before the first official recording in 1998 of a specimen in that area (see *Nature* **406**, 114; 2000).

The fake was spotted by *Nature* staff and confirmed by authors of the 1998 paper as being a manipulated version of their image.

Although the IRD has now cleared Séret and Pouyaud of faking the photograph, the institute has reprimanded them for failing to exercise the necessary rigour required in scientific research. Serre, a former consultant to the IRD, denies any wrong-doing. An enquiry by French police is continuing.



Cryptography on the front line

As the 'war on terrorism' unfolds, some politicians are calling for controls on the availability of encryption software. But many computer scientists claim such moves would play into the terrorists' hands. David Adam reports.



Crack team: the FBI is focused on terrorists — but could encrypted messages hamper its efforts?

Bankers, shoppers and other Internet users now have access to standards of privacy previously only available to the military. Off-the-shelf encryption software is effectively unbreakable — even by the massed computing power of organizations such as the US National Security Agency and the Federal Bureau of Investigation (FBI). Put that power in the hands of a terrorist network, and the potential for harm is all too obvious.

No surprise, then, that in the wake of the terrorist atrocities in New York and Washington, attention has focused on the ability of individuals to communicate securely over the Internet through encrypted e-mails. Although there is little evidence that those behind the attacks used such coded messages, some politicians are already calling for stronger controls on encryption software.

In a speech just days after the attacks, Republican Senator Judd Gregg of New Hampshire called for the US government to be given back-door access into all encryption

software. Britain's Foreign Secretary, Jack Straw, has also entered the fray, dismissing those who have fought against such moves in the past as "naïve". And on 6 October, the Dutch government announced that, as part of its counter-terrorism action plan, it intends to regulate the use of cryptography.

Coded warning

The events of 11 September had an immediate impact on public opinion — a poll conducted two days later indicated that 72% of Americans believed that anti-encryption laws would help to prevent repeat attacks.

But most experts on computer security argue that restrictions on encryption software would be expensive and impractical. Worse, they say that the net result would be to undermine the security of legitimate Internet users — rendering government and business more vulnerable to cyber-attack. But given the public statements of politicians such as Gregg and Straw, computer scientists are preparing for a reprise of the debate over

privacy and security that they thought they had won in the 1990s.

"We've been through these arguments before, but legislators seem to have short memories," says Bruce Schneier, chief technical officer at Counterpane Internet Security, a company based in Cupertino, California, that provides computer security services. "Limits on encryption and systems that ensure governments have access to encrypted messages will do little to thwart terrorist activities," he argues. "At the same time they will significantly reduce the security of our own critical infrastructure." (see Commentary, page 773.)

Encryption software uses mathematical algorithms both to scramble the contents of e-mails, by reordering the underlying data, and to decipher the encoded version. The algorithms are activated — and so protected — by numerical 'keys' typically containing 10 or more digits. One set of keys is widely circulated, and these are used to encrypt messages. But individual users also have private keys, which are used to decode messages. The algorithms

and their mathematical relationships with the keys are too complex for security agencies to crack, so access to the private key is in practice the only way to read an encoded message.

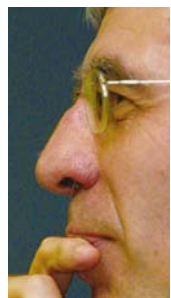
Intelligence and law-enforcement chiefs have long been concerned about the potential misuse of such programs. Indeed, former FBI director Louis Freeh in the late 1990s warned repeatedly that terrorists could be using encryption software to plan their actions, and he urged the US Congress to approve restrictions on its use and distribution.

Added restrictions

But Schneier claims it is impossible to limit the spread of cryptography. "Cryptography is mathematics and you can't ban mathematics," he says. There are almost 1,000 software products that use cryptography, available in more than 100 countries. "You would have to ban them in every country and even then it won't be enough, as any terrorist organization with a modicum of skill can write its own cryptography software," he says.

Blanket restrictions on the use of encryption might also impede the use of computers and the Internet in activities such as online

banking and shopping — which rely on encryption for security. A degree of disruption to e-commerce may seem a small price to pay for greater security, but cryptography systems also protect vital safety systems, such as the computers used in air-traffic control. "Restrictions are



Clampdown: Louis Freeh (right) and Jack Straw want tighter curbs.

not possible from a practical point of view," argues Matt Blaze, a principal research scientist with AT&T Laboratories in

Florham Park, New Jersey.

If governments cannot crack encrypted e-mails and they are unable to stop people using them, what options do they have? One is to force manufacturers to introduce 'back doors' into their encryption software, allowing the content of encrypted messages to be monitored routinely. This can be achieved by a system known as key escrow, in which copies of all private keys are handed over to a third party and can be accessed on demand by government security agencies.

The arguments for and against key escrow raged through the 1990s. Agencies such as the FBI argued that it would allow secure monitoring of communications with little disruption to normal Internet use. Civil-liberties groups campaigned against key escrow on privacy grounds, whereas computer scientists concentrated on practical flaws. Researchers in the field say that it is currently impossible to build a system that is secure enough to hold all of the private keys and guarantee that they could not be

accessed by those intent on committing fraud or wreaking cyber-havoc. Particularly daunting are the human factors — ensuring that individuals working for the key-holding organizations cannot be bribed or otherwise manipulated into releasing keys.

"It's all very well protecting bars of gold because at least you can see if they're gone in the morning," says Richard Clayton, who works in the computer security group at the University of Cambridge. "But when you're talking about lots of numbers hidden on behalf of people and you can't even tell if they've been stolen, then you're talking about needing a very secure system indeed. We're just not capable of building such systems." Schneier agrees: "Stockpiling keys in one place over an extended time period is a huge risk just waiting for attack or abuse."

Another problem with key escrow is that there is little commercial demand for encryption software that can be accessed at will by a third party — even in the name of national security. "It's not easy to demand that individuals use designated software," says Wenbo Mao, a researcher in the mathematics, cryptography and security group at



Hewlett Packard's UK laboratories in Bristol. "There is no market demand for it." Computer security experts are concerned that legislation enforcing key escrow would make legitimate computer users wary of using encryption technology — rendering their

systems more vulnerable to attack.

With little incentive for software manufacturers to develop reliable key-escrow technology, the task falls to government agencies, which traditionally have kept this kind of research classified. But this approach is a problem, argues Mao — users have low confidence in a product that has not been subjected to widespread attempts to crack its codes. Indeed, the US government in the mid-1990s abandoned attempts to introduce its own key-escrow scheme, based on a system known as 'Clipper', after Blaze at AT&T exposed flaws soon after it was released. "Government-certified systems developed behind closed doors would be a potential disaster," agrees Brian Gladman, a computer security consultant who formerly served as secure systems director with Britain's Ministry of Defence.

Computer scientists thought that they had won these arguments — but now the world has been thrown into conflict, they are not so sure. "If encryption is used in issues such as terrorism, and there is no legal way that law enforcement has access, then that has to be an issue," says a spokeswoman for the British government's National Criminal Intelligence Service.

Britain, in fact, last year passed a law that



Summing up: Bruce Schneier urges governments to focus on improving computer security.

computer security experts point to as an example of the sort of legislation that might be proposed elsewhere in the current climate. The Regulation of Investigatory Powers Act, championed by Straw when he was home secretary, gives police wide-ranging powers to intercept e-mail traffic, and also allows them to force individuals to surrender their private decryption keys. Refusing to comply, or revealing that you have been asked to surrender your keys, can be punished with up to two years' imprisonment.

Key questions

These powers have not yet been invoked, so the impact of the law cannot be assessed. One problem is that the police must first show that seized private keys can be held securely. The scale of security needed for this more limited number of keys — which would not make such a tempting target — is not the same as that required for a full key-escrow system. But developing an appropriate system is still not easy. The British government admits that practicalities remain to be worked out, but says that it hopes to implement the law by the end of the year.

Given this, many computer scientists argue that the focus should not be on restricting the use of encryption, but on encouraging the development of stronger security systems to protect computer infrastructure vital for national and economic security.

To this end, President George W. Bush on 9 October appointed Richard Clarke, a former member of the National Security Council, to the post of special White House adviser for cyberspace security. "America built cyberspace and now it must defend cyberspace," Clarke said, in accepting the position.

Clarke's position on cryptography remains unclear. But even if he doesn't reopen the debate on encryption, other politicians and officials are determined to do so. Computer scientists who oppose such moves, it seems, will be forced to do battle once again. ■

David Adam is a news and features writer for *Nature*.

Science sans frontières

Will the European Union's member states ever put the goal of continental cohesion in science ahead of their individual national interests? Quirin Schiermeier considers the prospects for creating a 'European Research Area'.

The road towards European integration is littered with false starts and missed deadlines. In 1970, a panel chaired by the prime minister of Luxembourg, Pierre Werner, proposed full monetary union among the members of the European Economic Community — as it was then called — by 1980. Yet more than three decades after his report, several European Union (EU) nations, including Britain, have yet to decide whether they want to adopt the euro, the EU's nascent single currency.

It is a similar story in science. In 1973, Germany's Ralf Dahrendorf, then European Commissioner for research, suggested the creation of a European scientific area, within which member governments would coordinate their national research policies. The proposal languished. In fact, when Philippe Busquin, the Belgian former physicist who now occupies the post once filled by Dahrendorf, last year launched his 'big idea', dubbed the European Research Area (ERA), there were few in Brussels with a sufficiently long memory to comment on its similarity to Dahrendorf's stillborn scheme.

Busquin's proposal, published in January 2000, noted the importance of scientific excellence to the EU's continued economic competitiveness. Today, Europe produces around one-third of the world's scientific knowledge. But by many measures of scientific output and productivity, Europe performs poorly compared with North America and Japan, and when it comes to spending on research, the gap is widening (see table, opposite).

Among other things, Busquin blamed fragmentation of resources and infrastruc-



Raising the flag: Philippe Busquin wants to create a coherent, coordinated science policy for Europe.

ture, a lack of intergovernmental coordination, and obstacles to mobility of researchers — the downside, perhaps, of the cultural and linguistic diversity across the EU's 15 member states. His document lamented the lack of a pan-European vision: "It cannot be said that there is today a European policy on research. National research policies and the Union policy overlap without forming a coherent whole."

Far-reaching ambitions

The ERA aims to create the desired cohesion. Its goals include the harmonization of patent law, support for large-scale infrastructure projects and existing centres of excellence, and the creation of new, 'virtual' centres using networking technology. It aims to integrate scientists from the central and eastern European nations that are currently queueing up to join the EU, while strengthening scientific links between the EU and the rest of the world. It also stresses the

importance of commissioning research to address important policy decisions that face the EU. Most importantly, Busquin aims to create a common European 'market' for science, technology and higher education, similar to that which already exists for goods and services, by boosting the pan-European mobility of researchers at all career stages, and by improving coordination between national research-funding programmes.

Dahrendorf's proposal similarly stressed the need to coordinate national research programmes. But whereas his idea soon disappeared from view, Busquin's ERA proposal has entered the vocabulary of Europe's leaders, who adopted it at the Lisbon EU summit in March last year. The ERA has become a political priority, according to the public statements of national governments. "The ERA comes at the right time and has our full backing," David Sainsbury, Britain's science minister, told a meeting of scientists and policy-makers held in Brussels last month to

discuss the ERA's implementation.

But as the travails of monetary union reveal, even Europe's highest priorities can fall foul of inertia and conflicting national interests. Twenty months on from Busquin's original proposal, it seems that the obstacles that stand in the way of a true pan-European science policy are bigger than had been thought. One problem is that there are no policy instruments to force national research agencies to collaborate. The original thinking behind the Common Agricultural Policy (CAP) may be largely discredited, but those who are familiar with the workings of the EU argue that its success in coordinating national farming policies cannot be doubted. The ERA, they note, lacks such a powerful legal mechanism.

Given that a scientific equivalent of the CAP is not on the EU's agenda, the fate of Busquin's ERA lies in the willingness of national officials and representatives of pan-European research organizations to turn their enthusiastic statements into action. However, doubts are beginning to emerge. "There seems to be an absence of political will," says Enric Banda, secretary-general of the Strasbourg-based European Science Foundation (ESF).

Meanwhile, the European Commission — more specifically, its five-year, rolling Framework programmes for research — remains the driving force. The sixth Framework, due to start in 2003, is intended to ignite the touch-paper on the ERA. The 17.5-billion-euro (US\$15.9-billion) programme is currently being discussed by the European parliament, and will also be on the agenda of next week's meeting of the EU council of research ministers from the member states in Luxembourg. Both bodies will have a say on its final form.

Economics of scale

The sixth Framework proposal turns its back on the proliferation of smaller, collaborative projects that have characterized its predecessors, giving priority to large-scale projects and research infrastructure (see *Nature* **410**, 4; 2001). Large-scale projects are divided into 'Networks of Excellence' — long-term multidisciplinary projects in fundamental science — and 'Integrated Projects' — efforts on a similar scale in applied and industrial R&D. The proposal also gives participants more autonomy in managing individual projects, and aims to integrate scientists from the 11 candidate member countries of central and eastern Europe.

To Busquin's chagrin, however, the proposal has been attacked on several fronts. Many scientists fear that the new rules will see funding divided among already wealthy research groups. "There will be no 'equal and unbiased access' to EU funding, as claimed," argues Martin Grabert, head of the Brussels-based liaison office of Germany's science organizations. "The scheme is tailor-made for 'old-boy networks' with plenty of experience and resources." In Poland, Slovenia and

International research bodies must pay more than lip-service to the goal of coordinating their activities.

several other candidate EU member states, these worries are particularly acute. Scientists in these countries fear that they could be excluded, rather than integrated, by the proposal (see *Nature* **411**, 512; 2001).

Precision planning?

The sixth Framework proposal has also been criticized for its vagueness. Although it identifies a series of 'priority areas', defining general scientific themes, the details of each will not be filled in until the European Commission has called for 'expression of interest' early next year. In response to previous complaints from the scientific community, the commission has decided this time not to define the precise scope of each component of the Framework programme until it has gained some input on the sort of projects that EU researchers wish to pursue.

Seasoned observers of EU research policy have some sympathy with Busquin's predicament — in attempting to consult with scientists, he has come under attack from the European Parliament and the member states. For example, Gerard Caudron, the French socialist member of the European Parliament who is coordinating its response to the Framework proposal, has suggested that the priority area that covers climate and other environmental research, dubbed 'sustainable development and planet changes', should be completely rewritten, while others must be "fleshed out and better structured".

Although the unrest that surrounds the sixth Framework does not augur well for the ERA, Busquin has always stressed that his vision could never be realized through the Framework system alone. Indeed, the current Framework accounts for only 5.4% of total public spending on R&D across the EU. For the ERA to become reality, national governments must address the problems that currently restrict free movement of researchers across national borders, and the various

national and international research bodies must pay more than lip-service to the goal of coordinating their activities.

Busquin has made removing obstacles to mobility a high priority. He wants a scientist to be able to relocate from, say, London to Milan with the same ease that an American researcher might move from New York to San Francisco. "I would like researchers in Europe to be free and ready to go wherever an envisaged project has the best chance of realization," says Busquin. The European Commission has been trying to foster greater mobility through the Framework programme for years. The proposed budget for the sixth Framework, for instance, contains 1.8 billion euros for 'human resources' projects, including grants for training courses and postdoctoral fellowships.

But such activities barely scratch the surface of the changes desired under the ERA. To move things along, the European Commission has devised a mobility strategy, which was published in June. This details a range of reforms, including the harmonization of social-security rights and mutual recognition of degrees and diplomas. It also encourages the creation of 'mobility centres' — local offices that will deal with legal and administrative problems — and a network of national ombudsmen, who will handle practical complaints from guest researchers.

Obstacle course

But all of this will have to be enacted and paid for by the EU's member nations. For several months, Busquin has been campaigning hard to prepare the ground, taking his case personally to senior officials in the member states and trying to convince them of the need to set up inter-ministerial meetings to solve the problems that currently restrict most European scientists to their country of birth.

Although there are few reliable statistics, it is clear that the obstacles to mobility are many. Moving from one country to another can leave researchers out of pocket. In the absence of bilateral or European social-security agreements, for instance, visiting scientists may be forced to pay for social benefits they cannot enjoy, lose pension rights, or risk double taxation (see *Nature* **407**, 427–429; 2000). In the case of researchers travelling on grants paid in euros to countries such as Britain and Sweden, which do not yet use the single currency,

Big three: how Europe, the United States and Japan shape up in research and innovation

	European Union	United States	Japan
Researchers per thousand workforce	5.28	8.08	9.26
R&D intensity (spending as % of GDP)	1.92	2.62	2.91
Growth of R&D intensity since 1995 (%)	3.03	5.55	4.13
US patents per million population per year	73	315	249
EU patents per million population per year	135	144	134
World market share of high-tech exports (%)	18.49	25.04	12.62

Source: European Commission



EMBL



EMBL



CERN

Success stories: is there scope for more projects modelled on the European Synchrotron Radiation Facility (top), the European Molecular Biology Laboratory (middle) and CERN (above)?

fluctuations in the euro's value can also cause problems (see *Nature* **406**, 551; 2000).

Even high-profile EU-funded projects can suffer. The Centre for Advanced Manufacturing Technologies at the Technical University of Wrocław in Poland, for instance, last year received a grant of 1.1 million euros, of which 300,000 euros were earmarked for fixed-term employment of scientists from elsewhere in Europe. An ideal candidate, who was enthusiastic to move, was soon found at the Technical University of Dresden in Germany. But Knut Grossmann, a mechanical engineer, is still in Dresden — the two universities were unable to find a proper contractual arrangement for his guest stay. A request for advice, which the Wrocław centre sent to the European Commission in Brussels, has so far remained unanswered.

There are also concerns — particularly acute in southern Europe — that scientists who spend time away from their home country will find themselves at the back of

the employment queue on their return. “Whoever moves loses his seat,” observes Ramon Marimon, Spanish secretary of state for science and technology policy, quoting a traditional Spanish proverb.

Some countries have made moves to improve the situation. Belgium, for instance, earlier this year approved a 2.5-million-euro programme designed to facilitate the involvement of Belgian scientists in EU programmes, which will include measures to encourage the mobility of researchers. But other changes on a national level threaten further restrictions. For instance, proposed regulations on the maximum time allowed between graduation and assistant professorship in Germany (see Correspondence, page 773) could restrict the ability of its young scientists to change labs or go abroad during their postdoctoral period.

European Commission officials have no illusions about the difficulty of the task ahead. “It will take a regular dialogue between the commission, national governments, regions, communities and universities to progressively remove the various obstacles,” says Elisabeth Colinet, head of a unit that is charged with implementing the ERA. “However, when it comes to legislative and administrative changes, the member states have to take the lead.”

Managing or meddling?

Busquin's goal of coordinating the activities of national research agencies is similarly beset with difficulties. Some observers even claim that this would run counter to one of the EU's fundamental principles — that of ‘subsidiarity’. This states that the EU should not carry out activities that can be performed adequately at a regional or national level. But given the complaints about the adequacy of the current set-up detailed in the original ERA proposal, commission officials reject this argument.

Irrespective of arguments about subsidiarity, however, national governments are anxious about the idea that tax money paid by their citizens could be spent abroad. Tony Mayer, the ESF's head of scientific coordination, says that thinking is still dominated by the notion of *juste retour*, in which participation in EU programmes is proportional to financial contributions. “All states try jealously to get back what they have paid in,” he says.

To break free from such limitations, Busquin promotes a flexible approach in which individual research agencies of the member states, with or without the involvement of the European Commission, are encouraged to come together to support projects of mutual interest. In the jargon of the ERA, such projects have ‘variable geometry’.

The idea is not a new one. The ESF, for instance, which was created as an association of national research agencies, has coordinated pan-European projects of this nature for many years, linking related activities in

different countries. Last year, it went further, launching its EUROCORES programmes, under which participating agencies agree to pool their funding and operate a centralized system for the selection of projects. EUROCORES programmes are each likely to have budgets of several million euros. Peer review of proposals for the first, on anthropology and linguistics, is now under way.

Framework for progress

Busquin wants to see more of this type of activity. But in the long run, many scientists and policy-makers believe that realizing the ERA may require the creation of an independent European funding agency, fed by funds from national organizations and complementing the EU Framework programme (see *Nature* **411**, 871; 2001).

The ESF is especially keen on the idea of creating a new European research council, along the lines of the US National Science Foundation, and is touting itself as the natural body to take the plan forward. “We are willing to play a role,” says the ESF's Banda. He adds that the ESF plans to appoint a high-level group to carry out an in-depth study of the potential advantages and disadvantages.

The Danish government, which will begin a six-month presidency of the EU in July, is also enthusiastic about the idea. But bringing a European research council to fruition will take years — and may ultimately fall foul of the subsidiarity principle.

Some scientists argue that the EU should instead concentrate on building from existing successful models of European scientific cooperation — centralized facilities such as the European Synchrotron Radiation Facility in Grenoble, the European Molecular Biology Laboratory in Heidelberg, and CERN, the European particle physics laboratory near Geneva. “If the EU is serious about increasing Europe's competitiveness, it should think about ways of setting up at least five new large facilities, preferably in the fields of genome research and cell and developmental biology,” says physiologist Brian Heap of the University of Nottingham, vice-president of Britain's Royal Society.

Ultimately, however, the fate of the ERA lies in the willingness of EU member states to turn their stated commitment to Busquin's vision into reality. “The ERA's makers will be judged by what they do, not by what they say,” says Marimon. “If they really act, the ERA could indeed become one of the main achievements in our century.” ■

Quirin Schiermeier is *Nature's* German correspondent.

European Research Area proposal

♦ europa.eu.int/comm/research/area.html

Sixth Framework proposal

♦ europa.eu.int/comm/research/nfp.html

European Commission proposal for mobility of researchers

♦ www.cordis.lu/rtd2002/fp-thematic/mobility.htm

European Science Foundation

♦ www.esf.org

Masai Mara tourism reveals partnership benefits

Wise use of income could satisfy local communities while aiding conservation efforts.

Sir—As government budgets for conservation in protected areas decline, there is increasing need for other mechanisms to fund conservation efforts: for example, private-sector involvement. Tourism has long been viewed as a potentially benign source of funding for protected areas, but relatively few sites worldwide have so far been able to generate significant resources. Even where they have, sustainable conservation will result only if the revenues generated are properly managed and allocated.

A case in point is the Masai Mara National Reserve in Kenya, world famous for its huge density and diversity of wildlife and an annual migration of more than a million wildebeest. Visitor entrance fees alone could generate \$US5.5 million annually for conservation and surrounding community development. This equates to earnings of over \$3,500 per square kilometre, some 12 times more than the estimated requirement for effective management¹.

To date, little revenue has been collected by the local councils overseeing the reserve, and very little has been reinvested. This all appears set to change as the Trans Mara County Council has just contracted the Mara Conservancy, a

Kenya-based private consortium, to manage its portion of the reserve, including ticketing, revenue collection, tourism management, security and wildlife conservation. In its first few weeks of operation in June and July, the Mara Conservancy generated more than five times the amount previously collected by the council over a whole year, as well as attracting additional donor funding². Already new security and management equipment has been purchased and staff are being paid on time.

Such a dramatic turnaround bodes well for a reserve that has suffered from uncontrolled and unmanaged tourism impact, and which has witnessed considerable declines in many of its wildlife species³. Moreover, this example could have far-reaching consequences for both wildlife and local Maasai communities. Much of the Masai Mara's wildlife is seasonally reliant on dispersal areas outside the reserve, which results in considerable conflict between people and wildlife. Local communities are entitled to receive 19% of reserve revenues as compensation, but have received little or no money since the mid-1990s. The turnaround in financial management in Trans Mara suggests that these

communities could soon see the benefits they are due, which should be equivalent to at least \$1 million around the whole reserve.

However, as revealed at workshops supported by the United Kingdom's Darwin Initiative, local residents remain suspicious of this new, private management structure, which appears as yet another neo-colonialist plot to expropriate their land. The onus is on the Mara Conservancy to ensure that benefits flow to neighbouring communities via the local council as rapidly as it is upgrading the resources of the reserve management itself.

If financial resources can be effectively captured and appropriately deployed, then real conservation gains could be made, both inside and outside the reserve. The Masai Mara could become a model for tourism-based, sustainable, integrated conservation and development through public-private partnerships.

M. J. Walpole & N. Leader-Williams

Durrell Institute of Conservation and Ecology, University of Kent, Canterbury CT2 7NS, Kent, UK

1. James, A. N., Gaston, K. J. & Balmford, A. *Nature* **401**, 323–324 (1999).

2. Mbaria, J. *East African* **354**, 8 (2001).

3. Otchicho, W. K., de Leeuw, J. & Prins, H. H. T. *African Journal of Ecology* **38**, 202–216 (2000).

Italian immunology well but hoping to do better

Sir—I was delighted to read, in the same issue, your News Feature on science in Italy (*Nature* **412**, 264–265; 2001) and your *Naturejobs* article on careers in clinical immunology (*Naturejobs* 19 July, 5; 2001). However, contrary to the statement in *Naturejobs*, clinical immunology and allergy is a medical specialism in Italy. There are 24 Italian university postgraduate training centres, accepting about 50 medical doctors a year who specialize in allergy and clinical immunology after a four-year programme. In addition, there are about 40 PhD programmes in basic and clinical immunology.

Basic and clinical immunology is relatively strong in Italian academic institutions. A recent survey of the productivity of Italian biochemical research commissioned by the Research and Education Ministry showed that immunological disciplines fare much better than other biomedical sciences. And an analysis by the US Institute for

Scientific Information showed that 20% of the 100 most productive Italian investigators are either basic or clinical immunologists (*Immunology Letters* **73**, S51, 2000).

I agree with your News Feature that immunology suffers from many of the problems common to Italian biomedical research: insufficient funding and little collaboration with industry. It is to be hoped that the new Research and Education Minister, Letizia Moratti, will recognize the strategic importance of investment in science.

Gianni Marone

President of the Italian Society of Allergology and Clinical Immunology (SIACI) and of the Italian Federation of Immunological Societies (IFIS), Clinical Immunology Division, University of Naples Federico II, 80131 Naples, Italy

Postdocs in France need less red tape, more cash

Sir—Having lived in France and worked at the Pasteur Institute for four years, I agree with your editorial “Mixed fortunes in France” (*Naturejobs* 16 August, 3; 2001)

that, in general, there is no lack of candidates for postdoctoral positions in European (or North American) institutions. I also agree that the definition of a ‘quality’ candidate varies depending on a lab director's priorities and vision.

Certain particularities of the French system explain why some organizations there say they have difficulties attracting quality candidates. Funding for newly graduated French PhDs makes it very difficult for them to stay in France to continue their training. There are very few external postdoctoral funding sources, so unless candidates are admitted to permanent positions with the CNRS or INSERM they cannot be paid from a lab's operating budget (as is done in the United States and Canada). Hence, many high-quality candidates are forced to leave France and continue their training elsewhere.

For foreign candidates, the situation in France is complicated. The lack of domestic funding for postdoc salaries means that most foreign candidates must arrive with their own funding. Postdocs from other European countries have an

easier time than those from elsewhere. Many of the former can obtain funding from organizations in their home countries or from other European community sources, and immigration issues are simplified.

For non-European citizens, however, the visa and immigration requirements and procedures are complex. Many institutions have little or no administrative structure to help foreign candidates navigate the complicated rules and regulations. In this regard, things are slowly improving in some institutions, such as the Pasteur Institute, but not in others, such as the Ecole Normale Supérieure, where candidates are left to their own devices.

The French government could greatly improve the pool of quality candidates available to its research organizations by developing new sources of funding available to all postdocs, and by simplifying the administrative procedures for foreigners wishing to train in France.

Andrew A. Munk

*New Ventures Bioconsulting,
1015 Strauss Street, Suite 201, Brossard,
Québec J4X 1T2, Canada*

Planned law could put German research at risk

Sir — The planned new university law in Germany (*Hochschulrahmengesetz*) is intended to make universities more competitive by abolishing the *Habilitation* requirement, and by creating new positions (*Juniorprofessoren*) to allow people to carry out independent research before attaining a tenured position (see News Feature, pages 768–770).

We fully support these aims, and firmly believe that the establishment of many more groups of independent young investigators is essential to making the German research system more flexible, attractive and productive. These groups could be modelled on the US system's assistant-professor grade, or on the independent research groups which are already established at several German universities and Max Planck institutes.

But the new law contains strict time constraints that severely undermine the competitiveness of German research, at least in the area of life sciences. We strongly believe this law must be changed before it goes through parliament.

The intention is that there will be no more than six years between the diploma (the examination required before beginning a PhD project) and the start of a

junior professorship, which itself is limited to a maximum of six years. In a manner that may seem typically German, both time limits will be strictly enforced. After a transitional period, the completion of a period as *Juniorprofessor* will become the prerequisite for securing a regular, tenured position.

As a consequence, all PhD and postdoctoral work will have to be completed within only six years, which is much too short for many biomedical and other natural-science projects. It will become unrealistic to complete an innovative PhD project and then move to a new lab for postdoctoral research. Up to now, almost all postdocs working both in Germany and abroad have taken more than six years to complete their work. Yet under the new law, any scientist exceeding the six-year limit will be excluded from a post as *Juniorprofessor*. Far from luring German postdocs working in foreign countries back home, the new law will actually exacerbate the brain drain.

The time limit will effectively abolish the current postdoctoral system in Germany. The prerequisite for a post as *Juniorprofessor* will be an excellent PhD thesis, with no requirement for additional scientific work.

This regulation will increase the dependence of PhD students on their advisers, a dependency that may be even worse than that experienced by more senior scientists doing their *Habilitation* (which the new law is intended to mitigate) and will render everyone's subsequent academic career contingent on the initial stage of their scientific activity. Almost all *Juniorprofessoren* will lack the invaluable experience of having done research in more than one lab (within Germany or abroad) before starting their own group, and will be less inclined to change research topics.

In sum, the strict adherence to this six-year rule will not only deprive German research labs of postdocs and exacerbate the brain drain, but will result in poorly qualified academics, lacking the valuable experience of having worked in different laboratories or on different subjects. Moreover, the emphasis on time, rather than quality, will encourage research projects that are safe and dull, rather than innovative yet risky.

More than 1,000 scientists (*Professoren*, group leaders, postdocs and students) support the views expressed here; their names and addresses are listed at www.zmnh.uni-hamburg.de/jentsch/unterschriften.

Thomas Jentsch

*Valentin Stein Zentrum für Molekulare Neurobiologie, University of Hamburg,
Falkenried 94, D-20246 Hamburg, Germany*

Hunger added an edge to Vincent's sharp wit

Sir — Glancing through old issues of *Nature* recently, I ran across the Millennium Essay "Sounds of Silence" by Robert C. Young (*Nature* **408**, 141; 2000). The picture of *Skull with a Burning Cigarette* by Vincent van Gogh is striking. It is worth a comment, for van Gogh is arguably the artist whose paintings are mostly easily recognized. *Skull with a Burning Cigarette* scarcely fits our preconceived notions of van Gogh's work.

It was painted in the winter of 1885–1886 in Antwerp, and the artist was "mocking the procedure in drawing classes, where a skeleton invariably served as the basis of anatomical studies" (I. F. Walther and R. Metzger *Vincent van Gogh, The Complete Paintings* 217; Taschen; 1993).

The artist was objecting to the lifelessness of this procedure, the very opposite of what he wished to bring out in his paintings. The cigarette was added as he had just taken up smoking to deaden the hunger he felt so often. It has been suggested that this painting may, in a sense, be regarded as Van Gogh's first self-portrait.

Ajoy K. Baksi

*Department of Geology and Geophysics,
Louisiana State University, E235 Howe Russell
Building, LSU Baton Rouge, Louisiana
70803-4101, USA*

Bacteria in biofilms are living an unnatural life

Sir — Jean-Marc Ghigo in his interesting Letter (*Nature* **412**, 442–445; 2001) states the oft-repeated canard that "most natural bacterial populations are found associated with environmental surfaces". This is not true.

Bacteria are the most numerous organisms in the biosphere. A quick look at a sample from the oceans, the largest habitat on Earth, would reveal individual bacterial cells (when stained for their DNA) scattered across the darkened field, like stars on a clear night.

Of course there are bacteria in biofilms, but in the biosphere these are the exception rather than the rule. Their more numerous free-living brethren may tell us more about why these bacteria live in such an unnatural state.

David L. Kirchman

*College of Marine Studies,
University of Delaware, 700 Pilottown Road,
Lewes, Delaware 19958, USA*

Protecting privacy and liberty

The events of 11 September offer a rare chance to rethink public security.

Bruce Schneier

Appalled by the events of 11 September, many Americans have declared so loudly that they are willing to give up civil liberties in the name of security that this trade-off seems to be a *fait accompli*. Article after article in the popular media debates the 'balance' of privacy and security — are various types of increase in security worth the consequent losses to privacy and civil liberty? Rarely do I see discussion about whether this linkage is valid.

Security and privacy are not two sides of an equation. This association is simplistic and largely fallacious. The best ways to increase security are not at the expense of privacy and liberty. Giving airline pilots firearms, reinforcing cockpit doors, better authentication of airport maintenance workers, armed air marshals travelling on flights and teaching flight attendants karate are all examples of suggested security measures that have no effect on individual privacy or liberties.

Security measures that reduce liberty are most often found when system designers fail to take security into account from the beginning. They're Band-Aids, and evidence of bad security planning. When security is designed into a system, it can work without forcing people to give up their freedom. Take, as an example, securing a room. Option one: convert the room into an impregnable vault. Option two: put locks on the door, bars on the windows and alarms on everything. Option three: don't secure the room; instead, post a guard to record and check the identity of everyone entering.

Option one is the best, but is unrealistic. No vault is impregnable, getting close would be extremely expensive, and turning a room into a vault greatly reduces its usefulness as a room. Option two is the realistic best, combining the strengths of prevention, detection and response to achieve resilient security. Option three is the worst, as it is far more expensive than option two, and the most invasive and easiest to defeat of all three options. It's also a sign of bad planning: designers built the room, and only then realized that they needed security. Rather than installing door locks and alarms, they take the quick way out and invade people's privacy.

A more complex example is Internet security. Preventive countermeasures help significantly to protect sites against 'script kiddies' but fail against smart attackers. Detection and response are key to providing security on the Internet. My company catches hackers all the time, by monitoring the audit logs of network products: firewalls,



IDSs, routers, servers and applications. We don't eavesdrop on legitimate users, read mail or otherwise invade privacy. We monitor data about data, and find abuse that way.

We detect yesterday's attacks by watching for their signatures, and tomorrow's by noticing and investigating anomalies. We can respond in time to thwart these attacks. This monitoring doesn't work automatically; it requires people to separate real attacks from false alarms, to investigate anomalies and to pursue attackers relentlessly. It's not perfect, but combined with preventive security products it is more effective, and more cost-effective, than anything else.

There are strong parallels between Internet security and the real world. All criminal investigations look at surveillance records. The lowest-tech version of this is questioning witnesses. In the current investigation, the FBI is looking at airport videotapes, airline passenger records, flight-school class records and financial records. The effectiveness of the investigation is directly related to the quality of the examination.

Some criminals and terrorists are copycats, who do what they've seen done before. To a large extent, this is what hastily implemented security measures try to prevent. But others invent new methods, as we saw on 11 September. We can build security to protect against yesterday's attacks, but we can't guarantee protection against tomorrow's: the hacker attack that hasn't been invented, or the terrorist attack still to be conceived.

Demands for even more surveillance miss the point. The problem is not obtaining data, it's deciding which are worth analysing and interpreting. Everyone leaves an audit trail through life; the FBI quickly pieced together

the terrorists' identities once it knew where to look. More data can even be counterproductive. The National Security Agency and the CIA have been criticized for relying too much on signals intelligence, and not enough on human intelligence. The East German police collected data on four million people, yet they did not foresee the overthrow of the government because they invested heavily in data collection instead of interpretation. We need more intelligence agents on the ground in the Middle East debating the Koran, not sitting in Washington arguing about wiretapping laws.

People are willing to give up liberties for vague promises of security because they think they have no choice. What they're not being told is that they can have both. It would require us to discard the easy answers. It would require designers to build security into systems from the beginning instead of tacking it on at the end. It would require the structuring of incentives to improve overall security rather than simply decreasing its costs. And it would make us all more secure.

Some broad surveillance, in limited circumstances, might be warranted as a temporary measure. But surveillance should not be designed into our electronic infrastructure. As the saying popularized by Thomas Jefferson goes: "Eternal vigilance is the price of liberty." Historically, liberties have always been a casualty of war, but a temporary casualty. This war — a war without a clear enemy or end condition — has the potential turn into a permanent state of society. We need to design our security accordingly. ■

Bruce Schneier is at Counterpane Internet Security, 19050 Pruneridge Ave, Cupertino, California 95014, USA. This is an edited version of an article in *Crypto-Gram* at www.counterpane.com.

DAVID NEWTON

Honest Jim: the sequel

Further misadventures of one of the most influential scientists of our day.

Genes, Girls and Gamow

by James D. Watson

Oxford University Press: 2001. 304 pp. £18.99

Horace Freeland Judson

When Atheneum published James Watson's *The Double Helix*, in the spring of 1968, it became a remarkable best seller. To many back then, the ostensible subject will have seemed recondite: a narrative of the elucidation of the structure of DNA, fifteen years earlier, by Watson and Francis Crick.

Jim, sceptical of the general reader's tolerance for science, left out all but the barest bones of the technical details. Otherwise the book had everything. It had style, rough, awkward, but all the more effective for that — reading it was like riding in a jeep over ploughed fields: jolting, but it got you there. It had minor players sharply caricatured. It had Francis, whose brilliance, sophistication, and success with women Jim admired and bitterly envied. It had a real live heroine: Rosalind Franklin, working on the structure of DNA at King's College London, dead by the time the book appeared, whom Jim portrayed in a way that has angered many women ever since. It had excitement: the competition Jim painted with Linus Pauling to get the structure (and never mind that the race against Pauling was factitious, the real competition being with Franklin, though not acknowledged). It had a strange central character: Jim presented himself as a *naïf*, a holy fool, the trickster of myths. Above all, the story had a prize, a discovery that in its explanatory power and its consequences has proved unique in the history of science, transcendent.

Over the published narrative hovered a vague, lascivious *frisson*, countless references to pretty girls, popsies, hints of sexual encounters, like helium balloons after a children's party, bouncing against the ceiling. The text went through at least two extensive revisions. I have never read the earlier drafts, but one knows much of what's in them. The opinions about other scientists were more pungent. (One observer, years ago, pointed out that Watson and Crick's "organs of contempt" were hypertrophied.) The ribbons on those balloons were tied to — well, one wrist or another. The narrative was full of scandalous titbits.

Watson circulated those early drafts widely among scientists and others who had been on the scene. The title then was "Honest Jim", a reference both to Kingsley Amis's novel *Lucky Jim*, published in 1953, and to the fact that some at King's College London resented Watson and Crick's getting the structure by using data that were not from



A life of incidents and encounters: the science, loves and more of Jim Watson.

their work but from Franklin's, which had come to them without her knowledge. Response to the drafts was outrage. Crick led the charge. Pauling and others followed on his heels. Jim revised and revised. Nonetheless, Harvard University Press, which had the book under contract, was told by the university's lawyers not to publish it. The book went to Atheneum, a relatively small but highly successful trade publisher.

Watson went on to publish other books: textbooks that made him richer, and more recently a collection of essays. He became a promoter and critic of science, perhaps the most influential public scientist of our day, in large part because he never seems to hesitate in saying exactly what he thinks, be it ever so scathing. Wherever Watson is, in the room or on the printed page, there's tension.

Now we have *Genes, Girls and Gamow* — a further memoir, but one that shares none of the qualities of *The Double Helix* except the jarring prose style. Watson flashes back to his time in Cambridge, incidentally restoring some of the scandalous bits cut from



The Double Helix. He then carries his autobiography forward three years, to the summer of 1956 when he is about to begin an assistant professorship at Harvard. A synoptic epilogue disposes of various people and discoveries up to 1968.

He does serve up one flamboyant character, George Gamow, who wrote to Watson and Crick in the summer of 1953 to urge on them the next step, the coding problem, or how the sequence of the four kinds of bases in DNA dictated the sequence of twenty kinds of amino acids that make up a protein chain. Gamow suggested an approach based on the physical geometry of DNA. He was a

physicist of deep originality, among whose papers was the one in 1946 that established what Fred Hoyle later christened the Big Bang. He signed himself "Geo." (pronounced "Joe"), and was hugely convivial, loving parties, card tricks, practical jokes — even that rare creature, the scientific practical joke.

Gamow's approach to coding was mistaken but stimulating. His enthusiasms — like the RNA Tie Club, enrolling his friends working on coding — drove the small community of interested scientists along. Half a dozen papers, several of them ingenious and original, were circulated around the Tie Club. Watson acknowledges this, yet presents Gamow chiefly as a clown and heavy drinker. It was Crick who became grand marshal of the attack on coding, but he hardly figures in this account.

The true protagonist is, of course, Watson himself, in his mid-twenties but embarrassingly adolescent in his obsessive, fumbling pursuit of women. The love of his life was Christa Mayr, daughter of the Harvard evolutionary biologist Ernst Mayr. Yet she is a shadowy figure here: he never conveys her inwardness. Perhaps he never could: he appears interested only in himself. By the end, she escapes him. On the way, we slog through

a heaping up of incidents and encounters, hectic in tone, cumulatively trivial. Interminably, entries take the form, "The arrival, finally, of a letter from Christa let me enjoy the first real party of the fall, held on a Saturday night in Alfred Tissières's rooms in King's [Cambridge]. My head was hazy when I had to climb afterwards into college."

Above all, though, his science, treated just as sketchily as in *The Double Helix*, is less interesting, far less important, and Jim's part not central. In the years of this book, he was messing about with the structure and function of RNA. But what we now call transfer RNAs emerged from a theoretical suggestion of Crick's in a Tie Club paper and from the work of biochemists, chiefly Paul Zamecnik.

Messenger RNA was discovered later, first by Sydney Brenner and François Jacob on a visit to Caltech, working with Matthew Meselson. In the epilogue, Watson arrogates chief credit for messenger RNA to work he did at Harvard with Tissières and others, but that paper was secondary in every sense. Indeed, when Brenner completed the paper with Jacob and Meselson, they sent a copy to Watson — who asked that they hold it back to be published simultaneously with the Harvard group's. They acceded, waiting

some three months — a courtesy Watson does not acknowledge here, and which would be unthinkable today.

It's well known in publishing and molecular circles that Watson has been hawking this manuscript around for years. What's inexplicable is that Oxford University Press has bothered to pick it up. Who would have thought Jim Watson could be tedious? ■

Horace Freeland Judson is at the Center for History of Recent Science, George Washington University, Washington DC 20052, USA.

The world within the skull

I of the Vortex: From Neurons to Self

by Rodolfo Llinás

MIT Press: 2001. 264 pp. \$24.95, £16.95

Ilya Farber

At first glance, it's hard to tell what Rodolfo Llinás' book is about. The title suggests that it's about the neural basis of the self, but the individual chapters focus on topics that seem only distantly related: the evolution of the eye, the origin and development of the motor system, emotion, language, prediction, single-celled organisms, and the question of whether the Internet qualifies as a collective mind. A quick flip through the book's figures yields sea creatures and brain structures, sheet music, snarling animals and people, Buzz Aldrin's footprint — an entertaining collection, to be sure, but could all this possibly fit together into any sort of coherent project?

The answer is yes, although Llinás leaves it to the reader to forge some of the important connections. The novelty of his approach is that he begins by treating the self as a system in its own right, with its own intrinsic structure and dynamics, rather than as a higher-order monitor for other (sensory or cognitive) systems. Once the system is defined in this way, its core functions — including prediction, decision-making and motor control — can be studied from a variety of perspectives, such as the subjective, the behavioural and the neurofunctional. Similarly, its core structures, including the brain's basal ganglia and the neural circuits of the thalamus and cerebral cortex, can be investigated using the methods of comparative anatomy and computational neuroscience.

On the question of the relation between self and world, Llinás' position is that the self doesn't perceive or represent the world so much as duplicate it. Drawing on the example of dreaming, he argues that the brain is capable of sustaining a coherent world-model that evolves in real time according to its own internal logic. During waking consciousness, this

Clock-watching at sea

Taken from an instruction manual for sailors written by the sixteenth-century French navigator Jacques de Vaulx, this 'nocturnal' enabled sailors to tell the time at night, when the astrolabes used during the day were of no help. The illustration is just one of the abundance of images to be found in *Celestial Treasury: From the Music of the Spheres to the Conquest of Space* by Marc Lachièze-Rey and Jean-Pierre Luminet (translated by Joe Laredo; Cambridge University Press, £40, \$59.95), which charts the development of our understanding of the Universe as reflected in literature, the fine arts and philosophy.





Grey area: is our 'sense of self' a prerequisite for the control of our actions?

model is modulated by the senses, but it persists even when sensory input is temporarily cut off: close your eyes and you'll still be keenly aware that there is a desk in front of you. As far as your inner model is concerned, the desk is still there; and as that inner model is used to generate motor commands, any movements you choose to make (such as shifting your chair) will be adjusted to take the desk's presence into account. The model can thus be thought of as having two functional states: while 'online' — actively synchronized with the external world — it can be used to predict what will happen next or what the likely outcomes of different actions will be, and while 'offline', it can model scenarios from memory or imagination, or generate the realistic (if disordered) worlds of our dreams.

For Llinás, motor control is at least as crucial a component of the self as perceptual awareness. The self is the locus of agency (one of the classical components of older theories of self), the choice and control of

action; and Llinás argues convincingly that the centralization of motor control produces the need for the organism to monitor and predict its own bodily state. The pervasive, intimate, moment-to-moment 'sense of self' that we all enjoy is thus to be understood not as the product of some evolutionary leap in cognitive or perceptual sophistication, but as a functional prerequisite for the deliberate control of action. Llinás drives this point home with the striking example of the 'sea squirt', a tiny marine animal that celebrates the transition from its swimming larval stage to its sessile adult stage by digesting its own primitive brain. (One can't help but imagine it saying to itself: "So this is intelligence? Pfft! What a waste of energy!")

A serious flaw of the book is that two central ideas remain vague and poorly explained. The first is Llinás' neural theory of consciousness, which, although a close relative of some other well-known models, is unique in focusing attention on a form of global synchrony of neuronal action that may have an essential role in determining which perceptions reach consciousness. Unfortunately, the details of this theory are distributed across a decade's worth of journal articles, and the discussion in the book is brief and conveys little of the richness of the model. Presumably, Llinás was concerned about alienating the general reader with too much detail, but it is unfortunate that he has missed the opportunity to provide a clear, unified explanation of an intriguing model that deserves wider attention.

The second under-explained concept is the 'fixed action pattern', a term used by ethologists to describe stereotyped innate behaviours (such as a baby gull pecking at the red spot on its mother's beak), but which Llinás generalizes to include some learned

behaviours and even inner, mental 'actions' such as emotions. This concept is clearly intended to play a central role in the book, but it is defined only in passing and its meaning seems to shift so radically with successive chapters that it is unclear whether Llinás himself knows exactly what he wants it to mean.

The book's frustrations are perhaps inextricable from its strengths — Llinás does warn us in the preface that his discussion will span eight orders of magnitude, from ion channels to humans. The result is an unfinished product, but an undeniably exciting one. Anyone interested in new ways of thinking about the organization of the brain would do well to give it a look. ■

Ilya Farber is in the Department of Philosophy, George Washington University, Washington DC 20036, USA.

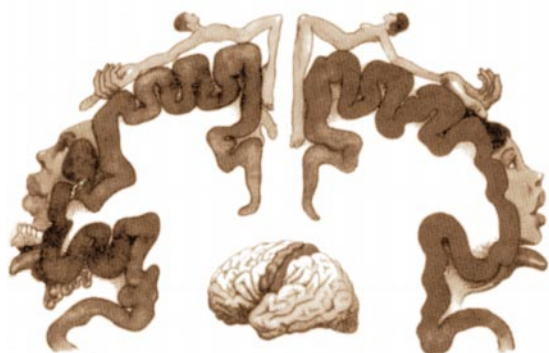
So why is it colder in winter?

Heavenly Errors: Misconceptions about the Real Nature of the Universe
by Neil F. Comins
Columbia University Press: 2001. 288 pp.
\$27.95, £18.95

Peter Coles

Astronomy is a curious subject to teach. Even the most unpromising fledgling scientist has probably, at some stage, looked at the night sky and wondered about the meaning of it all. Students usually therefore enter the classroom with some preconceived notions about astronomical matters. These notions are often naïve, sometimes inaccurate, and occasionally downright bogus. The teaching of astronomy does not, therefore, begin with a blank piece of paper, as it does with other topics in physical science, but with the correction of misconceptions that may be deeply held.

In *Heavenly Errors*, Neil F. Comins illustrates the ambivalent consequences of astronomy's peculiar allure with a series of commonly held misconceptions, misunderstandings and errors of logic. It is a promising idea for a book, particularly when the author has enlisted the willing help of thousands of undergraduate students to compile a list of frequently held wrong ideas about the Solar System and beyond. It is interesting to read of the origins of these misconceptions: Hollywood movies, astrology, the Internet and bad reporting of science all share some of the blame. But it's even more interesting to look at the different kinds of misconception and what they tell us about the chasm that often lies between scientific thinking and the 'common-sense' reasoning they represent.



Surgeon Wilder Penfield's maps showing how each part of the body is represented on two strips of the brain's cerebral cortex.

Ask why the weather is colder in the winter and you may well get the reply that, because its orbit is elliptical, the Earth is further from the Sun during winter than it is during summer and therefore receives less of the Sun's power at that time of year. This explanation fails to explain why the Southern Hemisphere experiences summer at the same time as the Northern Hemisphere experiences winter; that is, at the same stage of the Earth's orbit around the Sun. Talking through the logic of this example with students not only corrects the misconception, but also illustrates the scientific method by examining other necessary consequences of a given explanation before settling on the correct one. In this case, it is to do with the varying length of day and angle of the Sun in the sky.

Many of the examples presented by Comins are simple errors of fact. For example, "Polaris is the brightest star in the night sky", comes in at number 8 in the top 50 Cosmic Clangers (it is Sirius). Many others do not justify being called misconceptions at all. Time travel, which Comins takes to be self-evidently impossible, is not, as he claims, excluded by the general theory of relativity. On the other hand, he states that black holes are definitely not black because they give off Hawking radiation — this despite the fact that Hawking radiation has not yet been observed in an astronomical object.

And what is a misconception anyway? Contrary to popular belief, planetary orbits are not circular, and yet circles provide a useful approximate description for many purposes. We are told that they are actually elliptical, but this is itself an approximation that ignores perturbations from other bodies and relativistic effects. Most scientific explanations are misconceptions if you view them like this.

Much of the early part of *Heavenly Errors* is excellent, particularly its explanations of the basic astronomical properties of the Sun, planets and comets. But further on, the book goes badly off the rails. Through its conflation of fact and theory, and its blurring of the distinction between truth and approximation, it turns into a misguided crusade that encourages the rote learning of factoids as a means to "acquire a sound scientific foundation for understanding nature". I think this does more harm than good. T. H. Huxley, who knew a thing or two about science, put it best: "irrationally held truths may be more harmful than reasoned errors."

Peter Coles is at the School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG9 2HL, UK.

erratum In Edwin Taylor's review of *Mechanics of Motor Proteins and the Cytoskeleton* by Jonathon Howard (*Nature* **413**, 572; 2001), 'millimetre' in the third line of the fifth paragraph should have read 'micrometre'.

Science in culture

Reliquary and replication

A Genomic Portrait: Sir John Sulston by Marc Quinn.

Martin Kemp

We are all familiar with the stock portraits of scientists. They generally fall into three categories: evocative images of geniuses, in the Einstein manner; seated or standing subjects with suitable attributes, typically a piece of apparatus; and men and women "of action", posed as if working in the natural habitat of their labs.

Knowing what someone looks like may satisfy a basic human instinct, but it doesn't really tell us anything of importance about the scientist's significant intellectual and personal qualities.

A wholly alternative and headline-grabbing approach has been devised by the English artist Marc Quinn, who first came to fame for his sculpture of his own head cast in frozen blood. Commissioned to produce an image of Sir John Sulston, former director of Britain's Sanger Centre and international leader of the Human Genome Project, the artist and subject have collaborated to create something more like a religious reliquary than a standard likeness.

Sulston's own DNA was broken into random fragments and replicated in bacteria, which were cultured on a plate of agar jelly. The growth of the colonies was terminated at the point where they became visible as glassy beads scattered across the plate. The translucent, glistening 'picture', sharing its small scale with a snapshot rather than a grand painting, has been set within a broad silver frame, so shiny that the observer's own face is reflected in it.

But is it a 'portrait'? In that it contains certified fragments of the blueprint for a Sulston clone, it is in one sense a representation of him, or rather a re-presentation. It is as incontrovertibly individual as his fingerprints, containing those tiny quotients of genetic material that distinguish Sulston from Quinn, and all of us from primates. However, this level of individuality is not readily accessible to the spectator. It would need a Sulston and a Sanger Institute to undertake the analysis that would disclose the individualizing components. We simply can't tell, other than as an act of faith, that the beady traces embody Sulston's DNA rather than Quinn's.

I think it's not so much a functioning portrait as a kind of relic set in a precious vessel. In essence, it's akin to a bone fragment from Saint John the Baptist's finger, set in an incongruously elaborate masterpiece of the goldsmith's art — or, in a more secular context, the cherished curl of blond hair from a Victorian child preserved in a silver locket. There is that



Marc Quinn's 'portrait' of Sir John Sulston.

frisson of awareness that we are in the presence of something that is really part of the actual person rather than a mere optical record at one or more removes from the physical original. However, the traditional relic shares the problem of recognizability that bedevils Quinn's image.

The dusty fragment of ancient phalanx or fragile clipping of hair looks much like any old bone or curl of blond locks. Given the nature of Sulston's great achievement, the DNA reliquary could not be more appropriate, and it will clearly serve as an effective memento for visitors to the National Portrait Gallery in London. What it clearly does not do is serve as a wider exemplar for a new genre of individualizing portraiture. It is notable in this respect that the small exhibition staged to greet the portrait of Sulston contains coloured photographs of sitter and artist to tell us what they look like.

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

Sir John Sulston's 'portrait' will be on display at the National Portrait Gallery, London, until 10 February 2002.

Visualizations: *The Nature Book of Art and Science is a collection of essays edited by Martin Kemp (published by Oxford University Press and the University of California Press; £20, \$35).*

Matthew Cobb

779

Fundamental feelings

Antonio Damasio

The groundwork for the science of emotion was laid down most auspiciously over a century ago, but neuroscience has given the problem a resolute cold shoulder until recently. By the time that Charles Darwin had remarked on the continuity of emotional phenomena from non-human species to humans; William James had proposed an insightful mechanism for its production; Sigmund Freud had noted the central role of emotions in psychopathological states; and Charles Sherrington had begun the physiological investigation of the neural circuits involved in emotion, one might have expected neuroscience to be poised for an all-out attack on the problem. It is not usually appreciated that the probable cause of the neglect of the topic was the improper distinction between the concepts of emotion and feelings.

Some traits of feelings — their subjective nature, the fact that they are private, hidden from view, and often difficult to analyse — were projected onto emotions, so that they too were deemed subjective, private, hidden and elusive. Not surprisingly, neuroscientists were disinclined to give their best efforts to a problem that did not seem to be amenable to proper hypothesizing and measurement. Somewhat alarmingly, this conflation of the two concepts persists, as does the idea that the neurobiology of feelings is out of reach. A clarification is in order.

An emotion, be it happiness or sadness, embarrassment or pride, is a patterned collection of chemical and neural responses that is produced by the brain when it detects the presence of an emotionally competent

stimulus — an object or situation, for example. The processing of the stimulus may be conscious but it need not be, as the responses are engendered automatically. Emotional responses are a mode of reaction of brains that are prepared by evolution to respond to certain classes of objects and events with certain repertoires of action. Eventually, the brain associates other objects and events that occur in individual experience with those that are innately set to cause emotions, so that another set of emotionally competent stimuli arises.

The main target of the emotional responses is the body — the internal milieu, the viscera and the musculoskeletal system — but there are also targets within the brain itself, for example, monoaminergic nuclei in the brainstem tegmentum. The result of the body-targeting responses is the creation of an emotional state — involving adjustments in homeostatic balance — as well as the enactment of specific behaviours, such as freezing or fight-or-flight, and the production of particular facial expressions. The result of the brain-targeting responses is an alteration in the mode of brain operation during the emotional body adjustments, the consequence of which is, for example, a change in the attention accorded to stimuli.

Emotions allow organisms to cope successfully with objects and situations that are potentially dangerous or advantageous. They are just the most visible part of a huge edifice of undeliberated biological regulation that includes the homeostatic reactions that maintain metabolism; pain signalling; and drives such as hunger and thirst. Most emotional responses are directly observable either with the naked eye or with scientific probes such as psychophysiological and neurophysiological measurements and endocrine assays. Thus, emotions are not subjective, private, elusive or undefinable. Their neurobiology can be investigated objectively, not just in humans but in laboratory species, from *Drosophila* and *Aplysia* to rodents and non-human primates.

A working definition of feelings is a different matter. Feelings are the mental representation of the physiological changes that characterize emotions. Unlike emotions, which are scientifically public, feelings are indeed private, although no more subjective than any other aspect of the mind, for example my planning of this sentence, or the mental solving of a mathematical problem. Feelings are as amenable to scientific analysis as any other cognitive phenomenon, provided that appropriate methods are used. Moreover, because feelings are the direct consequences of emotions, the elucidation of emotional neurobiology opens the way to

Emotion

Emotion and feelings are closely related but separable phenomena; their elucidation, at long last, is now proceeding in earnest.

elucidating the neurobiology of feelings.

If emotions provide an immediate response to certain challenges and opportunities faced by an organism, the feeling of those emotions provides it with a mental alert. Feelings amplify the impact of a given situation, enhance learning, and increase the probability that comparable situations can be anticipated.

The neural systems that are involved in the production of emotions are being identified through studies of humans and other animals. Various structures, such as the amygdala and the ventromedial prefrontal cortices, trigger emotions by functioning as interfaces between the processing of emotionally competent stimuli and the execution of emotions. But the real executors of emotions are structures in the hypothalamus, in the basal forebrain (for example, the nucleus accumbens) and in the brainstem (for example, the nuclei in the periaqueductal grey). These are the structures that directly signal, chemically and neurally, to the body and brain targets at which alterations constitute an emotional state.

No less importantly, recent functional imaging studies reveal that body-sensing areas, such as the cortices in the insula, the second somatosensory region (S2) and the cingulate region of the brain, show a statistically significant pattern of activation or deactivation when normal individuals experience the emotions of sadness, happiness, fear and anger. Moreover, these patterns vary between different emotions. Those body-related patterns are tangible neural correlates of feelings, meaning that we know where to look further to unravel the remaining neurophysiological mysteries behind one of the most critical aspects of human experience. ■

Antonio Damasio is in the Department of Neurology, University of Iowa College of Medicine, 200 Hawkins Drive, Iowa City, Iowa 52242, USA.

FURTHER READING

- Damasio, A. R. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (Harcourt Brace, New York, 1999).
Davidson, R. J. & Irwin, W. *Trends Cogn. Neurosci.* **3**, 11–22 (1999).
Panksepp, J. *Affective Neuroscience: The Foundations of Human and Animal Emotions* (Oxford Univ. Press, New York, 1998).
Vuilleumier, P., Driver, J., Armony, J. & Dolan, R. J. *Neuron* **30**, 829–841 (2000).



Preserved for the afterlife

Sarah Wisseman

Ancient Egyptians used sophisticated combinations of natural substances to embalm the human body. Over time, they modified their recipes to balance quality of preservation with cost and availability of materials.

Over the years, scientific examinations of mummies have taught us a great deal about embalming practices, funerary rituals and burial economics in ancient Egypt. But that knowledge has often come at the cost of damaging the specimen. Mummies are an irreplaceable archaeological resource (Fig. 1), and these days the challenge is to extract new information without harming them.

On page 837 of this issue¹, Buckley and Evershed show how a great deal can be learned from virtually non-destructive analyses of mummies. They reveal that Egyptian embalmers used a greater variety of natural materials than previously reported, and that cheap and readily available plant oils and animal fats were probably used as bases for more costly and exotic resins.

The Egyptians believed that no one could enter the afterlife unless the most important part of the spirit, the *ka*, could return to the body. So the dead body had to be protected from decay and preserved in a recognizable and life-like form. After centuries of experimentation, embalmers learned that they could achieve excellent preservation by removing decay-causing organs (the intestines, liver, lungs and stomach), and by treating the body with substances, such as salts, resins, cedar oil, palm wine, myrrh, cassia, gum, honey and bitumen, that had drying or antimicrobial properties.

Embalms treated their clients according to what their families could afford: pharaohs were mummified with expensive mixtures and specially woven linen, but the poor received only cheap purges before being sent home to be wrapped in old clothes and used sheets. Depending on the wealth of the family and the customs of the time, mummies were buried in coffins (made of wood, fired clay or stone) or laid in communal pits. Both the materials and the various treatments used are described by ancient historians such as Herodotus² (fifth century BC), and are supported to some extent by modern scientific analyses.

At the height of their art (tenth century BC or twenty-first dynasty; Fig. 2), embalmers mummified bodies so well that scientists today can occasionally recover DNA from tissue and hair samples³. Unfortunately, however, many mummies never survived past the nineteenth century. If they escaped attack by tomb robbers and collectors, they



Figure 1 **Wrapped for eternity** — a mummy from the Roman period with a face portrait. This magnificent example, the Artemidorus mummy, is now in the British Museum, London.

were ground up for medicine, used as fuel, or destroyed in public unwrappings conducted as entertainment⁴. And although some twentieth-century studies^{5,6} delivered a huge amount of information, they involved not only unwrapping a mummy but full medical autopsies, virtually destroying the specimens.

More delicate analyses pose several difficulties for chemists: there are only a few well-preserved mummies of known provenance and date; museum curators are reluctant to permit sampling of material; and embalming fluids become degraded over time.

5000 BC Predynastic	Naturally desiccated bodies
2663 BC Old Kingdom	Plaster-wrapped mummies Evisceration introduced (fourth dynasty)
2066 BC Middle Kingdom	Cartonnage masks introduced
1549 BC New Kingdom	Removal of brain through the nose Lavish use of resins
1064 BC Third Intermediate	Height of embalmer's art (twenty-first dynasty)
332 BC Hellenistic	Embalming declines
30 BC Roman	Elaborate wrappings Face portraits

Figure 2 **Notable aspects of the embalmer's art listed by historical period.** 'Cartonnage' refers to the layers of papyrus, gum, fabric and plaster used to make masks and other fittings for mummies. The practice of mummification ended long after the advent of Christianity, sometime around the sixth century AD.

Buckley and Evershed¹ address some of these problems by using two complementary techniques based on a combination of gas chromatography and mass spectrometry. Their approach requires only tiny samples of tissue (less than 0.1 mg), and concentrates on diagnostic marker compounds that resist degradation. This essentially non-destructive approach conserves the artefacts and allows curators to comply with recent laws that require the respectful treatment of ancient human remains.

Previous chemical analyses of embalming substances confirmed the use of natron (a natural salt native to Egypt), juniper (rather than cedar) oil, camphor oil, myrrh, coniferous and other resins, galbanum (a gum resin) and beeswax^{7,8}. Bitumen, a petroleum derivative, has also been reported in a few cases⁹. Buckley and Evershed's analyses identify the use of resins from conifers and *Pistacia* (another type of tree), coniferous pitch, balsam (perhaps), plant oils, some animal fats and beeswax. Widely available

plant oils and animal fats were probably the base materials to which more expensive ingredients (resins, for example, some of which may have been imported from Lebanon) were added.

The authors point out that the amount of coniferous resins and beeswax increases in later mummies, implying that these substances became more important with time. They also stress the variety and compositional diversity of embalming materials. This finding is consistent with what other researchers have found: embalming materials and procedures vary more than expected, both within single dynasties and between historical periods¹⁰. This could be a result of economics (the cost and availability of materials), changing fashions, and/or the preferences of particular embalming guilds.

Buckley and Evershed tested 13 mummies of known provenance and date spanning approximately 2,300 years. However, only one or two mummies from each dynasty or period are represented (except for the Roman period, 30 BC to AD 395, for which four mummies were tested). Larger numbers of mummies from each period need to be analysed before we can conclude, for example, that Egyptian embalmers never used bitumen from the Dead Sea, or that bitumen was used only during the Roman period. (If this substance was used exclusively during Roman times, perhaps the reason was that it was cheaper or easier to transport under Roman rule.) Ideally, these mummies should be selected from those of known provenance and date to provide maximum information, but this is by no means easy — many mummies now residing in museums around the world are woefully deficient in documentation. In some cases, such as Roman mummies that can be dated fairly precisely by the style of the accompanying portraits, chemical analysis could contribute to our knowledge about embalming practices during a particular period without knowing the specific site of the mummy's origin.

As more mummies are tested, a larger database of embalming substances from all periods will help archaeologists explain variation in terms of changing economics and customs, as well as knowledge of material properties. For example, a preference for scented cedar oil over juniper oil in a particular period might indicate a choice based on fashion and cost, rather than preservative properties. The two coniferous oils are similar, but juniper oil was presumably cheaper to import from Lebanon because juniper was a more common tree than cedar. Wealthy Egyptians may have deliberately chosen the more expensive embalming fluid to impress family and friends, just as well-to-do people today select exotic woods and metal trims for their relatives' coffins. ■

Sarah Wisseman is in the Program on Ancient

Technologies and Archaeological Materials, University of Illinois, 116 Observatory, 901 South Mathews, Urbana, Illinois 61801, USA.

e-mail: wiscarc@uiuc.edu

1. Buckley, S. A. & Evershed, R. P. *Nature* **413**, 837–841 (2001).
2. Herodotus *The Histories* (trans. de Selincourt, A.) Book II, 160–161 (Penguin, London, 1954).
3. Pääbo, S. *J. Archaeol. Sci.* **12**, 411–417 (1985).
4. El Mahdy, C. *Mummies, Myth and Magic* 24–51 (Thames & Hudson, London, 1980).
5. David, A. R. (ed.) *The Manchester Museum Mummy Project*

(Manchester Museum, 1979).

6. Fleming, S., Fishman, B., O'Connor, D. & Silverman, D. *The Egyptian Mummy: Secrets and Science* (University Museum, Philadelphia, 1980).
7. Coughlin, E. A. *Paleopathol. News* **17**, 7–8 (1977).
8. Benson, G. G., Hemingway, S. R. & Leach, F. N. in *The Manchester Museum Mummy Project* (ed. David, A. R.) 119–131 (Manchester Museum, 1979).
9. Proefke, M. L. & Rinehart, K. J. *Am. Mass Spectrom. Soc.* **3**, 582–589 (1992).
10. Ikram, S. & Dodson, A. *The Mummy in Ancient Egypt: Equipping the Dead for Eternity* 109 (Thames & Hudson, London, 1998).

Molecular dynamics

Slick switching of X-rays

Ferenc Krausz and Christian Spielmann

The structures of crystals, from metals to proteins, have successfully been explored with X-rays. An ultrafast switch turns this idea around and uses a crystal to control the timing of X-ray pulses.

Since their discovery some 100 years ago by Wilhelm Röntgen¹, X-rays have found important uses in hospitals, laboratories and the exploration of space. They have proved particularly useful in investigations of the microscopic structure of matter². By scattering X-rays from small molecules, large biopolymers or macroscopic crystals, scientists can determine how the constituent atoms are arranged relative to each other. But obtaining fundamental information about the dynamics of molecular reactions is much harder, as it requires the motion of constituent atoms to be followed over interatomic distances. To achieve this, the X-rays must be switched on and off within a

time period short enough to 'freeze' atomic motion. This is an enormous challenge.

On page 825 of this issue, DeCamp *et al.*³ show how to control the proportion of X-rays transmitted through a crystal on a picosecond timescale (1 picosecond is 10^{-12} s). With this method it may be possible to develop a sub-picosecond X-ray switch, which would be fast enough to track dynamic changes in molecular structure during chemical and biochemical reactions.

Traditional X-ray structural analysis, although it cannot follow the progression of reactions, provides insight into how molecules work, and so is a key experimental technique for the chemical and life sciences.

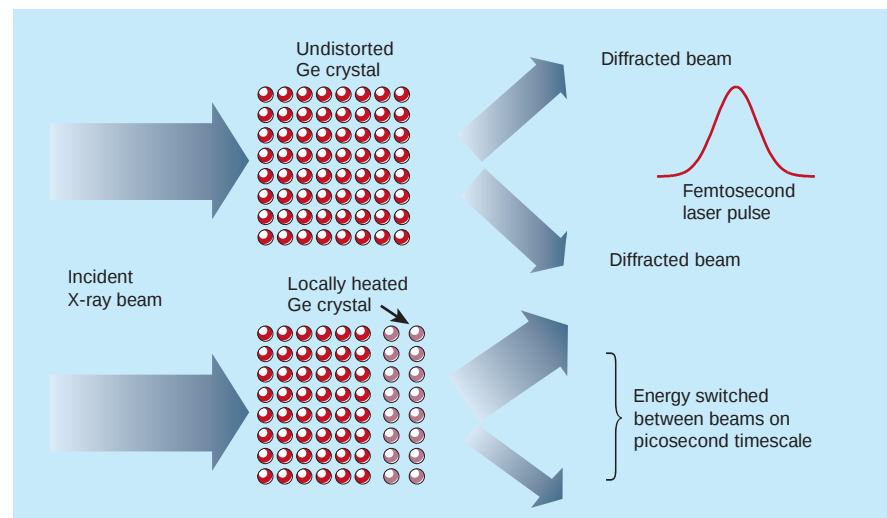


Figure 1 Blink and you'll miss it. An ultrafast switch for X-rays has been created by DeCamp *et al.*³ with the goal of generating ultrashort X-ray pulses that can track molecular dynamics during chemical and biochemical reactions. The switch is based on a germanium crystal that transmits X-rays with high efficiency (owing to the Borrmann effect). When the X-ray beam leaves the crystal it splits into two diffracted beams with roughly equal intensities. The structure of the crystal can be modified on a femtosecond timescale by using a laser pulse to heat and expand the lattice. The X-ray transmission is extremely sensitive to small lattice distortions, allowing the intensities of the diffracted beams to be modified on ultrafast (pico- to femtosecond) timescales, and even switched on and off, generating ultrashort pulses.

The equilibrium positions of atoms in complex molecules can be measured with a breathtaking accuracy of 10^{-13} m (about one-thousandth of a molecular bond length). Knowing the equilibrium structure of molecules allows physicists and chemists to predict how they might behave in different situations, but definite answers to many important questions require the direct observation of molecular dynamics.

Molecular processes, such as the breaking or formation of chemical bonds, have already been investigated by ultrafast pump-probe experiments, which use an intense femtosecond (10^{-15} s) laser pulse to trigger a reaction and a weak time-delayed probe pulse (operating at visible or longer wavelengths) to take snapshots of the subsequent dynamics⁴. But in these studies it is possible to follow changes in the atomic positions of only the simplest molecules. This is because the visible light can only probe the optical properties of weakly bound atomic electrons, from which the atomic positions can be inferred for simple molecules only. By contrast, 'hard' X-rays with wavelengths 5,000 times smaller than visible light can be absorbed or scattered by strongly bound 'core' electrons, providing direct information about the positions of the nuclei. So variation in the absorption or diffraction of hard X-rays can be unambiguously related to dynamic changes in molecular structure regardless of its complexity.

These prospects triggered a worldwide effort to develop sources of ultrashort pulses of hard X-rays. One successful approach has been to strip electrons from atoms (forming a plasma of ions and electrons) and then accelerate the electrons to velocities close to the speed of light using a powerful femtosecond laser pulse. When the energetic electrons re-collide with the atomic cores they produce a short X-ray burst at wavelengths characteristic of the atoms in the target. Such laser-driven sources of hard X-rays have already been used in proof-of-principle demonstrations of ultrafast X-ray absorption and diffraction measurements⁵⁻⁷.

Hard X-rays can also be created in a particle accelerator known as a synchrotron without the need for any collisions. Synchrotron radiation is emitted by high-speed electrons following a circular path through a strong magnetic field, and delivers pulses typically tens of picoseconds in duration. Last year, physicists produced sub-picosecond bursts of hard X-rays in a synchrotron for the first time by manipulating the electrons radiating the X-rays with a powerful femtosecond laser⁸.

The work of DeCamp *et al.*³ opens an entirely new chapter in controlling the time structure of hard X-rays. The authors modified the transmission of a synchrotron X-ray beam through a germanium crystal on a picosecond timescale by stimulating changes

in the crystal lattice with an ultrashort laser pulse. This X-ray 'switch' allows them to modulate hard X-ray beams regardless of their source of emission, and thereby turn the X-rays on and off to generate sequences of pulses, or even shaped pulses, over a wide range of hard-X-ray wavelengths.

The prototype X-ray switch developed by DeCamp *et al.* consists of a thin piece of germanium crystal irradiated by strong femtosecond laser light. By cutting and aligning the crystal in the correct way the authors have created an unusually high transmittivity for the incident hard X-rays. At the exit face of the crystal the transmitted X-ray beam is split into two diffracted beams, which propagate with nearly equal intensities in slightly different directions (Fig. 1). The relative transmittivity and intensity of the two X-ray beams is sensitive to minor distortions in the crystal structure. The exit face of the crystal can be heated up quickly with a femtosecond laser pulse; the heated volume expands, shifting atoms out of their equilibrium position in the lattice (by a process known as acoustic phonon excitation). So the structure of the lattice is perturbed, modifying the transmission of the incident X-ray beam and redistributing energy between the two outgoing X-ray beams. In this way the beams can be switched on and off, or the relative strengths of the two beams can be quickly altered.

When the X-ray switch is in operation the crystal atoms move only a fraction away from their equilibrium position, but this is still sufficient to switch a substantial fraction of X-ray energy from one beam to the other. This experiment beautifully demonstrates the sensitivity of X-ray diffraction to atomic positions, a feature that X-ray structural analysis itself relies upon. The X-ray energy flow into the outgoing diffracted beams is switched within the time it takes the atoms to leave their equilibrium position. This limits the minimum switching time to picoseconds in the current experiment, but the transmission of the X-ray beam can also be affected by small perturbations in the distribution of electrons around the atoms in the crystal lattice. Because electrons are much lighter than the nuclei forming the lattice, such perturbations (referred to as optical phonons) could be generated on a much faster sub-picosecond timescale with a sufficiently short laser pulse. Using electronic instead of acoustic perturbations could bring the switching time down to less than a picosecond, paving the way to eventually creating hard X-ray pulses of femtosecond duration.

How does this method compare with other techniques used to control X-rays on an ultrafast timescale? Laser-driven hard X-ray sources can generate pulses that are shorter than a picosecond but, because they emit photons in all directions, only a fraction of the X-ray photons can be focused on the



100 YEARS AGO

Paris was greatly excited on Saturday last when M. Santos Dumont, with his seventh balloon, successfully rounded the Eiffel Tower and returned to the shed at St. Cloud, thirty seconds within the thirty minutes allotted by the Committee of the Deutsch Prize. At the time of the voyage the wind, according to the *Times* correspondent, was blowing at the rate of twelve of thirteen miles an hour. At one period the balloon, travelling at the rate of thirty miles an hour, appeared as though it would collide with the Tower; the aeronaut, however, was able to control its movements without any apparent difficulty, and, as has been said, the journey was accomplished within the time limit agreed upon. M. Santos Dumont is to be congratulated upon the success which has at last attended the untiring efforts put forward by him towards the solution of the problem of aerial navigation.

From *Nature* 24 October 1901.

50 YEARS AGO

The second International Congress on Astronautics was held in London during September 3-8. It was attended by nearly fifty delegates representing societies and groups interested in astronautics from Argentina, Austria, France, Germany, Great Britain, Italy, Spain, Sweden, Switzerland and the United States... The latter part of the London Congress was devoted to a symposium of papers on the general theme of orbital vehicles, their construction and uses. There was general agreement among the speakers that such vehicles are possible from an engineering point of view; the first instrument-carrying vehicles could be built within ten to fifteen years; but man-carrying artificial satellites appear to be much further in the future... The greatest problem of interplanetary flight is that of propulsion, and in his paper "Interplanetary Travel between Satellite Orbits", Prof. Lyman Spitzer discussed a method of applying nuclear energy. It was suggested that an electrically accelerated ion beam could be used for achieving a gas ejection velocity of 100 km./sec. without the use of very high temperatures in the propellant gases. Such a unit could not be built with a large enough thrust/weight ratio to allow it to take off from the surface of a planet. It would be capable of travelling from a close-orbital station about the earth to a similar orbit about any other planet.

From *Nature* 27 October 1951.

material of interest. Synchrotrons are able to produce strong, collimated (laser-like) beams of hard X-rays, exposing targets to many more photons, but techniques that shape the electron beam in a synchrotron to generate ultrashort pulses require expensive modifications to the X-ray source. This is feasible only for dedicated beam lines. By contrast, DeCamp and colleagues' X-ray switch is a versatile tool that could be added to nearly every beam line without having to touch the source. Many challenges remain, such as shortening the switching time by using electronic excitations, and improving the switching efficiency to produce X-ray pulses with good contrast. But once these problems have been solved, such ultrafast switches

could become a key component in the X-ray toolbox for probing the structural dynamics of matter.

Ferenc Krausz and Christian Spielmann are at the Centre for Advanced Light Sources, Photonics Institute, Vienna University of Technology, Gusshausstrasse 27, A-1040 Wien, Austria. e-mails: krausz@tuwien.ac.at
spielmann@tuwien.ac.at

1. Röntgen, W. K. *Nature* **53**, 274–276 (1896).
2. Stout, E. G. & Jensen, L. H. *X-Ray Structure Determination, a Practical Guide* (Macmillan, New York, 1968).
3. DeCamp, M. F. *et al. Nature* **413**, 825–828 (2001).
4. Bhattacharjee, Y. *Nature* **412**, 474–476 (2001).
5. Raksi, F. *et al. J. Chem. Phys.* **104**, 6066–6069 (1996).
6. Rischel, C. *et al. Nature* **390**, 490–492 (1997).
7. Rose-Petruck, C. *et al. Nature* **398**, 310–312 (1999).
8. Schoenlein, R. W. *et al. Science* **287**, 2237–2240 (2000).

Developmental biology

Solving a sticky problem

Christof Niehrs

The Wnt family of proteins is a large one. Various combinations of its members control all sorts of developmental processes, including, it now seems, a pathway involving cell adhesion.

In this age of the genome, molecular biologists are often told that, to stay competitive, they must use genomics and *in silico* techniques when studying protein function. Well, here is a riddle for advocates of this approach. Proteins in the Wnt family are involved in cell-to-cell signalling during nearly every process one can think of in animal development, from the formation of the embryonic axes to kidney development, and from invertebrates to humans. In humans alone there are at least 19 genes encoding Wnt proteins, and ten receptors for these proteins.

So, how do you investigate the specificity of biological responses to Wnts? The answer, according to Winklbauer *et al.* (page 856 of this issue¹), is to use a cell-biological approach and — even more disturbingly — old-fashioned amphibian embryos. In this way the authors provide good evidence that vertebrates have three signalling cascades triggered by Wnts, which separately regulate cell differentiation, cell polarity and cell adhesion.

The best-characterized signalling cascade triggered by Wnt proteins is the canonical, or Wnt/ β -catenin, pathway (Fig. 1a). Details of this cascade first emerged from genetic analyses of fruitflies (*Drosophila*), where it functions in developmental processes such as the patterning of body segments and appendages. Since then we have learned, through work in vertebrates as well, that the canonical pathway is actually a large network. At its heart is the multifunctional protein β -catenin, most of which is bound to the internal face of the plasma membrane.

In the absence of signalling from Wnt, a large molecular complex ensures that β -catenin is rapidly targeted to the cellular protein-degrading apparatus. Activation of the pathway by a Wnt protein results in β -catenin being stabilized; it then enters the nucleus, where, with proteins of the high-mobility-

group family, it activates the expression of specific genes to control cell fate².

The existence of the planar cell polarity (PCP) pathway was also revealed by genetic analysis of *Drosophila*³. Cuticle cells in adult fruitflies secrete hairs, which are polarized by the PCP pathway so that they all point in one direction. The PCP pathway is also at work in vertebrates during gastrulation — the massive rearrangement of cells that produces the three main tissue layers, endoderm, mesoderm and epidermis, in the early embryo. During gastrulation, migrating cells become polarized by the PCP pathway and extrude lamellipodia — extensions that help movement — along one axis only⁴. The common theme to both processes is the polarization of the cytoskeleton.

In *Drosophila*, the receptor protein that lies at the start of the PCP signalling pathway is Frizzled. It is not known how this receptor is switched on in the fruitfly — that is, which protein binds to it to activate the pathway. But the vertebrate PCP pathway is triggered by Wnt11 (refs 5, 6; Fig. 1b), so a Wnt protein may well be involved in the *Drosophila* pathway, too. Instead of using β -catenin, the PCP pathway works through small GTP-binding proteins from the Cdc42/Rho family, which activate the transcription factor Jun.

Hints that there might be a third Wnt-triggered pathway came from the discovery^{7–9} that Wnt5a, together with another member of the Frizzled family, Frizzled-2, mobilizes Ca^{2+} ions within cells and thereby activates certain Ca^{2+} -dependent enzymes,

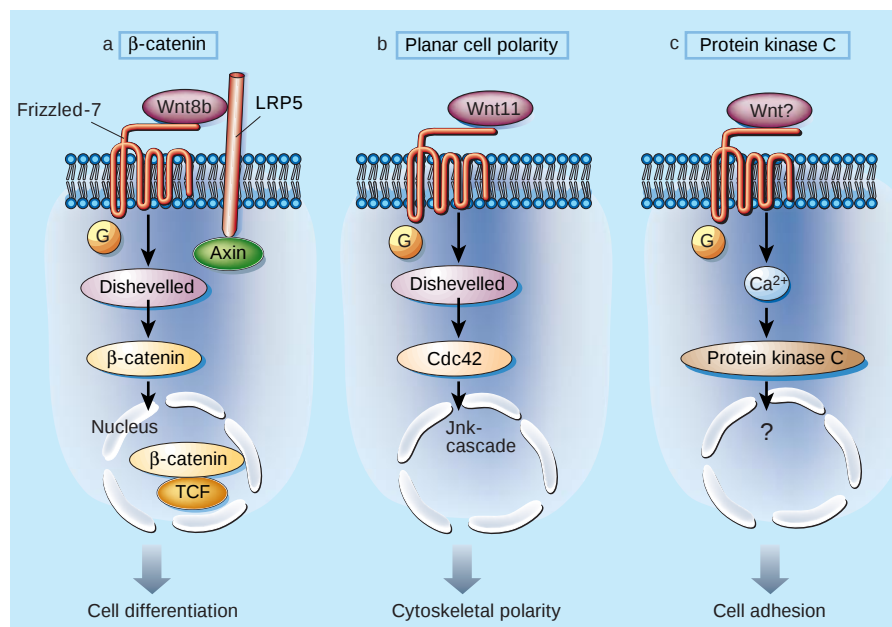


Figure 1 Three signalling cascades that are triggered by Wnt proteins in vertebrates. The choice of pathway activated by the Wnt receptor Frizzled-7 depends on which particular Wnt protein is outside the cell. a, Wnt8b triggers the β -catenin cascade^{10,11}; it achieves specificity by using a co-receptor, LRP5, which itself controls Axin¹³, a negative regulator of β -catenin. b, Wnt11 activates the planar cell polarity pathway^{4–6,12}. c, An unknown Wnt works through Frizzled-7 to induce an increase in the Ca^{2+} level inside cells and thereby activate protein kinase C^{1,7–9}. All three pathways also use molecular adaptors known as heterotrimeric G proteins (represented by G).

including protein kinase C (PKC). But Wnt11 (from the vertebrate PCP pathway) and Wnt5a belong to the same subclass and have similar effects when overexpressed, so it was unclear whether the activation of PKC actually represents a third pathway or is part of the PCP cascade.

Enter Winklbauer *et al.*¹, who provide strong evidence that this pathway is indeed distinct from the PCP cascade, and furthermore that it has a role in cell adhesion. Working with early frog (*Xenopus*) embryos, the authors studied the function of the receptor Frizzled-7 by inhibiting its production using specific antisense oligonucleotides. These molecules led to gastrulation defects, because the mesoderm and ectoderm failed to separate properly. To study the adhesive properties of the mesodermal cells, Winklbauer *et al.* used an ingeniously simple adhesion assay: they placed a ball of mesodermal cells on a piece of ectoderm and watched to see whether the mesoderm sank or remained where it was. Normal mesoderm stayed put, indicating that ectoderm and mesoderm have distinct adhesive properties. But mesoderm injected with anti-Frizzled-7 oligonucleotides sank in. This effect was reversed by injection of PKC but not by components of the β -catenin or PCP pathways, nor by Dishevelled, which functions in both cascades. Moreover, expression of Frizzled-7, but not of Frizzled-3 or Frizzled-8, could promote sorting of ectodermal cells.

So it seems that an as yet unknown Wnt protein and Frizzled-7 couple to a PKC-activating cascade that is distinct from the β -catenin and PCP pathways (Fig. 1c). However, these three pathways do have something in common: they are all sensitive to pertussis toxin, which blocks signalling through molecular adaptors known as heterotrimeric G proteins. What, then, determines which pathway is used? One clue lies in the particular combination of Wnt and Frizzled proteins. For example, Frizzled-7 will activate β -catenin when bound to Wnt8b (refs 10, 11), the PCP pathway with Wnt11 (ref. 12), and PKC in response to an unidentified Wnt.

How do Wnt–Frizzled combinations translate into specificity of pathway activation? For the canonical pathway, the answer involves the putative Wnt receptors LRP5 and LRP6, which function only in this cascade¹³. Just a few Wnt–Frizzled combinations can couple to LRP5/6, determining whether or not the β -catenin pathway is used. The existence of three Wnt pathways might also explain why there are so many secreted Wnt inhibitors, including sFRPs, Cerberus, Dickkopf and WIF: each inhibitor might be specific to certain Wnts or pathways. For example, Dickkopf1 blocks the β -catenin pathway by acting as an antagonist of LRP5/6 (ref. 13).

The next step will be to match other Wnt–Frizzled combinations to the three

pathways — that is, to decipher the Wnt–Frizzled code. Simple assays will be invaluable for this purpose, and Winklbauer *et al.*'s elegant study highlights one path to the answers. ■

Christof Niehrs is in the Division of Molecular Embryology, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

e-mail: niehrs@dkfz-heidelberg.de

1. Winklbauer, R., Medina, A., Swain, R. K. & Steinbeisser, H. *Nature* **413**, 856–860 (2001).
2. Wodarz, A. & Nusse, R. *Annu. Rev. Cell Dev. Biol.* **14**, 59–88 (1998).

3. Mlodzik, M. *EMBO J.* **18**, 6873–6879 (1999).
4. Wallingford, J. B. *et al.* *Nature* **405**, 81–85 (2000).
5. Tada, M. & Smith, J. C. *Development* **127**, 2227–2238 (2000).
6. Heisenberg, C. P. *et al.* *Nature* **405**, 76–81 (2000).
7. Slusarski, D. C., Corces, V. G. & Moon, R. T. *Nature* **390**, 410–413 (1997).
8. Sheldahl, L. C., Park, M., Malbon, C. C. & Moon, R. T. *Curr. Biol.* **9**, 695–698 (1999).
9. Kühl, M., Sheldahl, L. C., Malbon, C. C. & Moon, R. T. *J. Biol. Chem.* **275**, 12701–12711 (2000).
10. Medina, A., Reintsch, W. & Steinbeisser, H. *Mech. Dev.* **92**, 227–237 (2000).
11. Sumanas, S., Strege, P., Heasman, J. & Ekker, S. C. *Development* **127**, 1981–1990 (2000).
12. Djiane, A., Riou, J., Umbhauer, M., Boucaut, J. & Shi, D. *Development* **127**, 3091–3100 (2000).
13. Zorn, A. M. *Curr. Biol.* **11**, R592–R595 (2001).

Condensed-matter physics

Magnetic spins that last for ever

Piers Coleman

Phase transitions occurring at a temperature of absolute zero — quantum phase transitions — are hard to measure. A new type of quantum phase transition is now predicted to occur in two dimensions, rather than three.

Matter transforms itself into new ordered states through phase transitions. At a first-order phase transition, such as boiling or melting, the new phase emerges quickly following a sharp transition. At a second-order phase transition, the new order emerges gradually within the material. As the material passes into the ordered state, it passes through a remarkable stage known as a critical point, at which interactions with infinite range develop within the material. Surprisingly, the theory of critical phenomena, through a triumph of late-twentieth-century physics¹, did not involve any quantum mechanics. It turns out that any temperature above absolute zero (-273°C) destroys the coherence effects of quantum mechanics over the long distances where critical phenomena play a role. But quantum phase transitions have been observed in careful experiments, and on page 804 of this issue Si *et al.*² propose a new class of quantum phase transition that appear to completely reverse the situation that exists at a classical critical point.

About 30 years ago, physicist John Hertz³ first predicted the existence of a quantum phase transition in metals. Hertz proposed that when the temperature at which a magnetic phase transition takes place in a metal is reduced to absolute zero, a 'quantum critical point' would develop. In a conventional metal, interactions between electrons occur over a short range and the motions of individual electrons are largely independent. At a quantum critical point these interactions can grow until they have infinite range (Fig. 1a), but they also become infinitely retarded in time, so that the influence of a moving electron on another is felt long after it has passed. This causes the motions of electrons to become highly correlated, so they move

collectively and can no longer be considered as independent particles⁴.

One of the important side effects of the infinite-range interactions generated between electrons at a quantum phase transition appears to be the development of unconventional superconductivity — according to at least one school of thought, this is the origin of high-temperature superconductivity. Understanding the nature of these 'strongly correlated' states of matter is a pressing challenge in condensed-matter physics.

In the past decade, physicists have learned how to fine-tune magnetic materials so as to observe a quantum critical point⁵. Typically, this involves taking a ferromagnet (whose magnetic moments are parallel) or an antiferromagnet (whose magnetic moments are antiparallel), and lowering the critical ordering temperature to absolute zero by applying high pressures or by doping with suitable chemicals.

Two years ago, Schroeder *et al.*⁶ used neutron scattering as a probe to measure for the first time the energy distribution of the magnetic correlations at a quantum critical point. The material they studied, CeCu₆ doped with 10% gold, contains a dense array of magnetic cerium atoms, which behave as tiny quantum moments or 'spins'. At the quantum critical point these quantum spins develop antiferromagnetic correlations. Schroeder *et al.*⁶ found that the critical excitations seen in their neutron-scattering experiments appear to have a wide range of momentum values. This was unexpected, because it implies (following the uncertainty principle of quantum mechanics) that the underlying critical excitations are highly localized in space. The uncertainty principle says that knowing the precise position of a particle prevents you from knowing its precise momentum, and

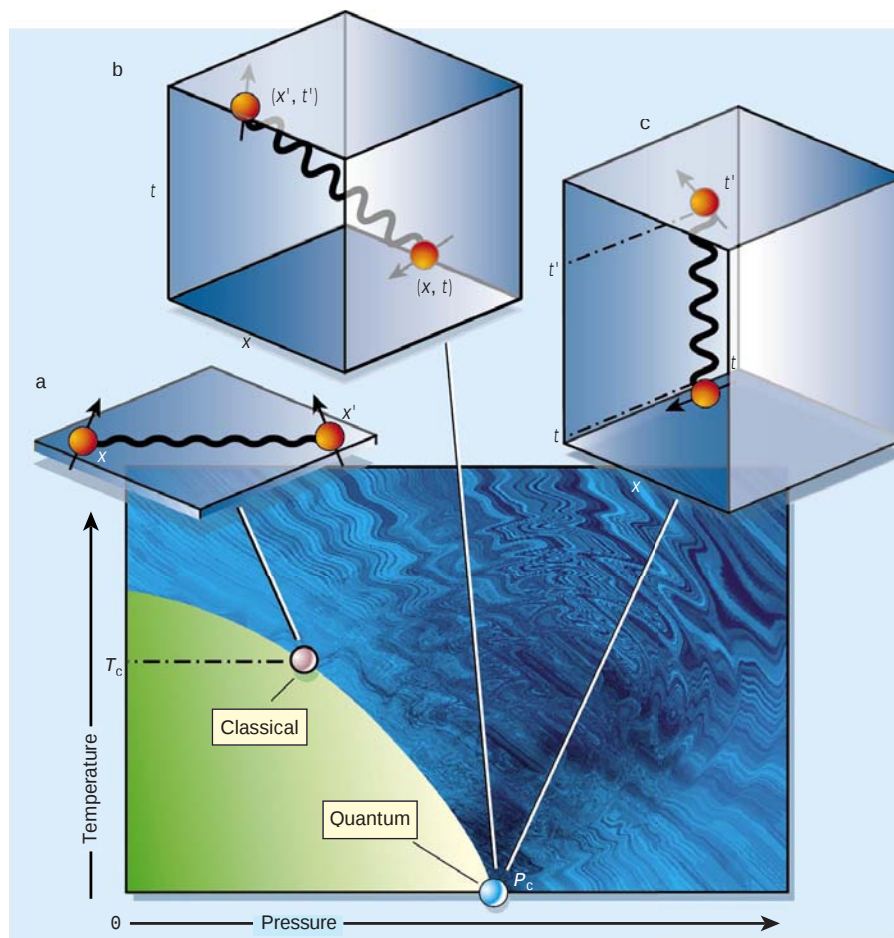


Figure 1 Classical and quantum phase transitions. Bottom, phase diagram of a magnetic metal, showing classical and quantum critical points. The transition temperature T_c vanishes at the quantum critical point. In the dark blue region above the quantum critical point, electrons may be strongly correlated and no longer move as independent particles. a, At a classical critical point, interactions between magnetic spins have infinite range in space. b, In the Hertz view of a quantum critical point, interactions between spins have infinite range in both space and time. c, At the new type of quantum critical point proposed by Si *et al.*², the interactions have infinite range in time, but are localized in space — this is the exact opposite to what happens at a classical critical point.

vice versa. The problem is that Hertz's theory predicted that the magnetic excitations should form a spatially delocalized wave, not a localized magnetic fluctuation.

Inspired by these measurements, Si and colleagues² have now proposed a new class of quantum phase transition in which the most singular correlations between the magnetic spins at the quantum critical point have an infinite range in time, but a finite range in space. This is the opposite to what happens at a classical critical point, and leads to magnetic interactions that are almost frozen in time yet localized in space. Si *et al.* propose that the interactions between the magnetic spins at different atom sites can be averaged, replacing the atoms surrounding any given site by an effective medium in which the magnetization fluctuates with infinitely long correlations in time.

This type of approach, called dynamical mean-field theory⁷, has become a powerful tool for analysing electronic materials with strong local correlations among the electrons. Using this approach, Si *et al.* examined the behaviour of a single spin inside this effective magnetic medium, and then used this to recalculate the local magnetic correlations of the effective medium. Self-consistency between the two parts of the calculation is achieved when the assumed magnetic correlations of the medium match up with those that result from calculating the behaviour of a spin within it.

Two possibilities emerge. If the magnetic medium is three-dimensional in character, Si and colleagues find that their theory corresponds to Hertz's theory, in which the critical spin fluctuations form a spatially delocalized wave. By contrast, the new class of quantum critical point emerges if they assume that the spin fluid around the local moment is two-dimensional. Many aspects of the two-dimensional solution are consistent with the observed properties of doped CeCu_6 — in particular, the theory of Si *et al.* can explain why the only energy scale in the neutron data is temperature itself.

But how can this be? How could a manifestly three-dimensional material exhibit two-dimensional spin fluctuations? To answer this question, Si *et al.* cite an observation made several years ago by Stockert *et al.*⁸, who noted that the critical behaviour of scattered neutrons is characterized by anisotropic momentum values, which suggests that the underlying magnetic fluctuations are quasi-two-dimensional. Si *et al.* point out that the arguments of Stockert *et al.*, taken together with their new theory, result in a complete picture of the critical magnetic behaviour in CeCu_6 .

Si and colleagues' theory will no doubt prove controversial, but it can be verified both theoretically and experimentally. On the theory side, some important aspects of the self-consistency of the calculations require further scrutiny. There are also many possible

experimental tests. For example, YbRh_2Si_2 has similar magnetic and thermodynamic properties⁹ to doped CeCu_6 ; if the theory of Si *et al.* applies generally then we can expect YbRh_2Si_2 to have quasi-two-dimensional features in its critical magnetic behaviour. Future neutron-scattering measurements will be able to check this prediction. A second test is suggested by cubic CeIn_3 , which undergoes a quantum phase transition at 2.5 gigapascals of pressure¹⁰. This system cannot display quasi-two-dimensional spin excitations because it is completely cubic. Does this material conform to the predictions of Hertz's theory, as expected on the basis of Si *et al.*'s results? This is an important question for future experiments.

Whether or not the theory of Si *et al.* passes all these tests, it has surely widened the debate and is an important step in the emerging discussion about quantum criticality that is likely to continue in the years ahead. ■

Piers Coleman is in the Department of Physics and Astronomy, Rutgers University, Frelinghuysen Road, Piscataway, New Jersey 08854-8019, USA. e-mail: coleman@physics.rutgers.edu

1. Wilson, K. A. *Rev. Mod. Phys.* **55**, 583–600 (1983).
2. Si, Q., Rabello, S., Ingersent, K. & Llewellyn Smith, J. *Nature* **413**, 804–808 (2001).
3. Hertz, J. A. *Phys. Rev. B* **14**, 1165 (1976).
4. Varma, C. M., Nussinov, Z. & van Saarloos, W. <http://xxx.lanl.gov/abs/cond-mat/0103393>
5. Stewart, G. *Rev. Mod. Phys.* **73**, 797 (2001).
6. Schröder, A. *et al. Nature* **407**, 351–355 (2000).
7. Georges, A. *et al. Rev. Mod. Phys.* **68**, 13–125 (1996).
8. Stockert, O. *et al. Phys. Rev. Lett.* **80**, 5627–5630 (1998).
9. Trovarelli, O. *et al. Phys. Rev. Lett.* **85**, 626–629 (2000).
10. Mathur, N. D. *et al. Nature* **394**, 39–43 (1998).

Archaeology

Those elusive Neanderthals

Chris Stringer and William Davies

The 'how, where and when' of possible Neanderthal coexistence with Cro-Magnons, and their extinction, continue to exercise a varied community of researchers. The latest interpretations of the fossil and archaeological records were aired at two meetings.

Will we ever really know what happened to the Neanderthals — the distinctive humans who occupied Europe immediately before the modern-looking Cro-Magnons appeared¹? At two meetings* held in late summer in Gibraltar and Liège, Belgium, there was plenty of new evidence from the fossil and archaeological records to discuss. But we are still far from pinning down the processes that caused Neanderthals to vanish about 30,000 years ago, after existing for at least 200,000 years before that.

Previous views on the question have ranged from a denial that they disappeared at all (they simply evolved into modern humans) to arguments that they were killed off rapidly by invading, vastly superior Cro-Magnons. These days, both DNA and morphological studies support the majority view that Neanderthals were indeed a separate lineage, and probably species, to modern humans. But equally there is growing evidence from dating techniques that Neanderthals did not vanish overnight, and that in some ways they were as behaviourally sophisticated as the Cro-Magnons.

In Gibraltar there was the prospect of some dramatic news, in two presentations by E. Trinkaus (Washington Univ., St Louis). In one he argued that modern features such as femoral shape and muscle attachments to the humerus of a late Neanderthal skeleton from St Césaire, France, dated to about 36,000 years ago, are morphological signs of behavioural change. These changes might reflect new patterns of activity, or the use of technological innovations. In the other, Trinkaus described his interpretation of a child's skeleton from Lagar Velho, Portugal, as showing a mixture of Neanderthal and Cro-Magnon features, implying that there was gene flow between the populations. These interpretations provoked discussion — not least because, in the first instance, he used skeletal plasticity as the explanation of a mixed morphology, and in the second he invoked gene flow between the two types. Several participants argued that the Lagar Velho skeleton is actually that of a modern human. But further details of this controversial specimen will be

published shortly and allow a more informed debate about it.

Increased knowledge of the vagaries of Europe's climate over the past 100,000 years is beginning to influence thinking about Neanderthal extinction. Cores from the Greenland icecap and the floor of the North Atlantic reveal remarkable oscillations in temperatures during that time. Figure 1 shows a vegetation curve based on pollen from the sediments of Lake Monticchio², Italy — the dramatic fluctuations show how severe the effects of the changing climate could have been on both Neanderthals and Cro-Magnons in reducing the environment's capacity to support them. According to one view (C. Finlayson, Gibraltar Museum), Neanderthals became extinct simply because they could not cope with the increasing severity of climatic change around 30,000 years ago. The Cro-Magnons, who according to Finlayson were better adapted to the increasing open-country environments, simply colonized the vacant habitats. An alternative view (proposed by R. White, New York Univ., and ourselves, among others) is that extinction probably stemmed from various factors, including climatic instability coupled with competition for resources from Cro-Magnons.

Neanderthals and Cro-Magnons are known to have occupied sites across Europe and western Asia. At the Gibraltar meeting, however, the archaeological presentations

centred on the Iberian peninsula. New ¹⁴C dates are available for one site, at Bajondillo in southern Iberia (A. Baldomero Navarro, Málaga Univ.). These seem to indicate a Cro-Magnon presence there about 34,000 years ago. This is slightly earlier than previously known for the region, but is still relatively late compared with elsewhere in Europe, and is consistent with other evidence that Neanderthals may have excluded the Cro-Magnons for a longer period in southern Iberia.

The diversity and complexity of human responses to their physical and social environments was a recurring theme. One source of evidence is the inferred arrangement of Neanderthal activities around hearths at the Catalan site of Abric Romaní (M. Vaquero, Univ. Rovira i Virgili, Tarragona), which could show that the social structure and organization of Neanderthals were — at least in part and in some places — similar to those of modern hunter-gatherers. At Abric Romaní there are indications that a variety of domestic activities took place around large hearths inside the rock shelter, with more specialized actions occurring at smaller hearths outside.

At the Liège congress there were three topics relevant to Neanderthal extinction. The first was an assessment of 'acculturation', meaning the influence Neanderthals and Cro-Magnons may have exerted on each other. Second, there was an assessment of what we mean by Aurignacian culture, and what it might tell us about modern human behaviour. The Aurignacian is the earliest Eurasian archaeological industry generally associated with Cro-Magnons, and is known from artefacts such as beads, bone carvings and certain stone tools, including small ones probably designed to be set compositely into a wooden shaft.

The third topic was assessment of the timing of the transition, 45,000–30,000

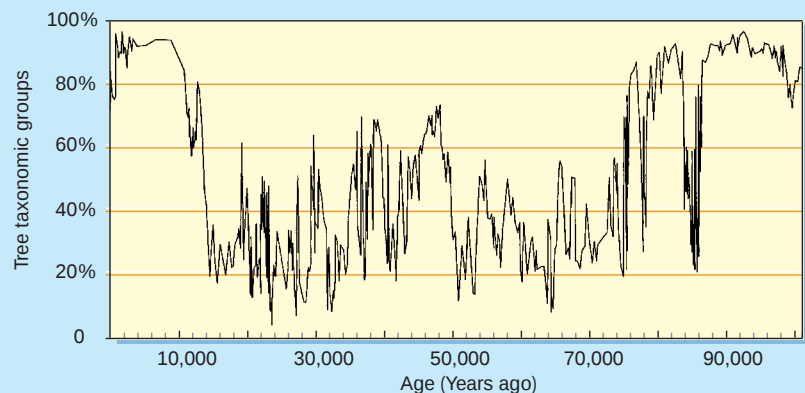


Figure 1 Climatic ups and downs in Europe. Here the number of tree taxonomic groups (expressed as a percentage) is used as the proxy indicator of variation in climate; the record was created from pollen in sediments of Lake Monticchio, Italy². Around the time of the Neanderthals' demise, about 30,000 years ago, there seem to have been very swift climatic shifts, occurring over intervals as brief as 20 years. But what contribution these made to Neanderthal extinction remains a matter of debate.

*Neanderthals and Modern Humans in Late Pleistocene Eurasia, Gibraltar, 16–19 August 2001; XIVth Congress Union Internationale des Sciences Préhistoriques et Protohistoriques (UISPP), Liège, 2–8 September 2001.

years ago, between the Middle Palaeolithic (Neanderthals) and Upper Palaeolithic (generally Cro-Magnons) in Eurasia. The Châtelperronian — which is characterized by a form of stone knife with a blunted diagonal edge, found for example at St Césaire — seems to be the only early Upper Palaeolithic industry that can definitely be associated with Neanderthals. But there are several other industries without associated fossil material.

Speculation about the effects that incoming Cro-Magnons might have had on Neanderthals in Europe depends on accurate dating of the archaeological cultures. Two schools of thought compete here. One holds that Neanderthals began the transition to 'modern' behaviour before Cro-Magnons arrived as, for instance, at Abric Romani. The other is that the arrival of Cro-Magnons with behaviourally modern Aurignacian technology stimulated a similar development in Neanderthals. To tackle these issues, archaeologists will have to assess how much significance can be attached to constructs such as the Middle Palaeolithic, Aurignacian and Châtelperronian.

A better chronology of behavioural change and Neanderthal extinction is also required. At the moment, archaeological ^{14}C determinations often come up with a younger age for bones than for charcoal — this is because bones contain proportionately less carbon, and are more susceptible to contamination with younger ^{14}C . Where both bone and charcoal from the same level have been dated, some bone dates statistically match those on charcoal, but others are much younger and often make little stratigraphic sense. Most Châtelperronian dates are on bone samples, and so could either be about the right age or be underestimates — without comparable charcoal samples we cannot be certain. Resolving this issue will have a great influence on thinking about when and how Neanderthals achieved 'modern', Upper Palaeolithic behaviour.

The ^{14}C determinations also need to be corrected to account for variability in ^{14}C production. Intervals between ^{14}C dates will be expanded or contracted when we convert them to calendar years, and such changes can alter long-established views on the chronological pace and even the succession of human behavioural change. Cross-calibration curves are one answer, and it has become possible to calibrate dates against a curve ('CalPal³') derived from a marine core⁴ and lake sequence⁵ (B. Weninger, Köln Univ.; O. Jöris, Monrepos, Neuwied).

These curves are working models only, but several speakers in Liège described how they calibrated their ^{14}C determinations with CalPal. For instance, not only have Middle and Upper Palaeolithic ^{14}C dates from France and Iberia been corrected to

calendar years, but they have also been subdivided into dates for bone and charcoal (O. Jöris). The charcoal dates are consistently older than those for bone, and the pattern was even clearer after calibration. Such techniques will no doubt develop apace over the next few years, but these curves are complex and progress will not be straightforward. We can be sure, given the discrepancy between ^{14}C and calendar years of up to four millennia or so during the period 40,000–30,000 years ago, that many radiocarbon ages in the phase when the last Neanderthals and the first Cro-Magnons are assumed to have coexisted will be underestimates. But we cannot yet tell whether the upshot will be a longer or shorter inferred period of coexistence.

What of the future for these kinds of studies? More reliable chronologies are one item on the wish list. Another is identifying the makers — Neanderthal or Cro-Magnon — of some enigmatic European industries in the range 30,000–45,000 years ago, such as the Szeletian and the Bohunician. These industries are characterized by stone 'leaf-points', which are probably a form of projectile tip. They have previously been considered to be a central and eastern European analogue to the Châtelperronian — that is, a Neanderthal behavioural foray into the 'modern' world of the Upper Palaeolithic⁶. But in rethinking the transitional period we might in some cases consider the possibility that they are Cro-Magnon. A final wish is to be able to extract DNA from the earliest Cro-Magnons and compare it with the growing number of Neanderthal sequences. This has proved difficult so far, both because of preservation problems and because it will be far trickier to tell whether Cro-Magnon, as opposed to Neanderthal, DNA is contaminated with our own DNA.

Above all, both conferences showed that European prehistory cannot be studied in isolation — we will be better able to answer many questions about Neanderthals and Cro-Magnons when we have a comparably rich fossil and archaeological record from western Asia. ■

Chris Stringer is in the Human Origins Programme, Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK.
e-mail: cbs@nhm.ac.uk

William Davies is in the Centre for the Archaeology of Human Origins, Department of Archaeology, University of Southampton, Avenue Campus, Southampton SO17 1BJ, UK.
e-mail: swgd@soton.ac.uk

1. Stringer, C. & Gamble, C. *In Search of the Neanderthals* (Thames & Hudson, London, 1993).
2. Allen, J. R. M. *et al.* *Nature* **400**, 740–743 (1999).
3. CalPal; <http://www.calpal.de>
4. Voelker, A. H. L. *et al.* *Radiocarbon* **42**, 437–452 (2000).
5. Kitagawa, H. & van der Plicht, J. *Radiocarbon* **40**, 505–515 (1998).
6. Allsworth-Jones, P. *The Szeletian* (Clarendon, Oxford, 1986).

Daedalus

The meaning of junk

The reading of the human genome has set Daedalus wondering what it all means. One puzzle is junk DNA, which is replicated down the generations as faithfully as useful DNA, but is useless or lethal if translated into protein. One theory is that it is a parasitic passenger. Daedalus thinks that nature is not so easily fooled; for him, junk DNA is a store of unused genes. He would like to think it might be a store laid in for the future. But more soberly, he agrees that it could be a relic, a cache of genes that came in handy in the past.

Human junk DNA should therefore tell us about the protohumans so vital to theory. How, for example, did our greatly extended childhood evolve? Presumably to accommodate the mighty human invention of 'unlearning' so many animal instincts: abandoning the fear, lust, rage and greed of animals and replacing them by the shrewd rational cunning of the true human. This evolution of new software is the hallmark of humanity, and should be written in our junk DNA, if only we can read it.

So DREADCO biochemists are studying the junk DNA revealed by the recent scanning of the human genome, looking for signs of the past. It should also throw up other markers — the development of the hand with its opposed thumb, or the foot and its bipedal gait, or the language instinct.

The DREADCO team are also looking for junk DNA laid down for future use. Even the most ardent believer in creative evolution must agree that most of this will have no value. In the same way, we need to store about a hundred times as much junk, or information, as we will ever use; we just don't know which 1% will be vital. But somewhere in among the junk the team hopes to read the things we should be preparing now: higher elasticity for the cardiovascular system, a better balance in women between pelvic shape and large brain size in babies, better-protected main joints, and a self-repairing brain.

But the biggest theoretical challenge is the 'panspermia' theory of Hoyle and Wickramasinghe. If indeed intelligent life is passed from one planet to the next by interstellar bacteria, those bacteria must carry in their junk DNA just the genes needed to develop such life. By contrast, mere earthly bacteria should have junk DNA reflecting their undistinguished past and unambitious future. Bacteria that colonize space will feel great evolutionary pressure to develop the ideal junk DNA for the job, very easy to spot.

David Jones

Motor-cortical activity in tetraplegics

It may eventually be feasible to reconstruct voluntary motor activity in the paralysed.

Paralysed patients may benefit from the development of an implantable brain–computer interface device that can bypass damaged motor pathways^{1–3}. But it is unclear whether chronically de-efferented areas will still be sufficiently excitable to respond to motor attempts^{4,5} if the motor cortex has been extensively reorganized^{6,7}, and, if they are, whether this excitability is somatotopically organized^{8,9}. Here we use functional magnetic resonance imaging to study brain activity in subjects with spinal-cord injuries while they are executing, or attempting to execute, movements of different limbs. We show that their motor-cortical activation closely follows normal somatotopic organization in the primary and non-primary sensorimotor areas. Our results indicate that any reorganization of the motor system that does occur in these patients does not affect attempt-related activation, and that it should be possible to access voluntary control signals by using a cortical neuroprosthetic.

We examined five subjects with spinal-cord damage (C5–C6 trauma-related injuries, 1–5 years after injury; mean age, 28.4 yr) who were asked to repetitively move, or attempt to move, their hands (adduction–abduction of the fingers), elbows (flexion), feet (toe flexion), left knee (extension) and lips (pursing) during alternating 30-s blocks. Subjects were instructed to attempt movements they were unable to make, rather than to imagine them, and they reported being able to follow these instructions. Their ability to execute the different movements was evaluated visually and by using surface electromyographic measurements.

One subject had residual use of his hands, whereas the other four, when attempting to abduct and adduct their fingers, elicited fasciculations of muscles in the wrist and palm. All five were unable to move their legs voluntarily. To assess the somatotopy of cortical activation, we used surface projections of statistical parametric maps¹⁰. Group comparisons were made with a healthy control group ($n=5$; mean age, 28.0 yr) by using surface renderings of 'group brains' (group averages of spatially normalized brains). As the pre- and post-central gyri of the individual brains extensively overlapped, they were also visible in the group surface renderings, including a 'knob' that marked the hand motor representation¹¹.

Figure 1 shows that activation was significant in the injured group when attempting hand movements (four subjects) on the two banks of the contralateral central sulcus (primary motor and sensory cortices), extending towards the pre- and post-central sulci (pre-

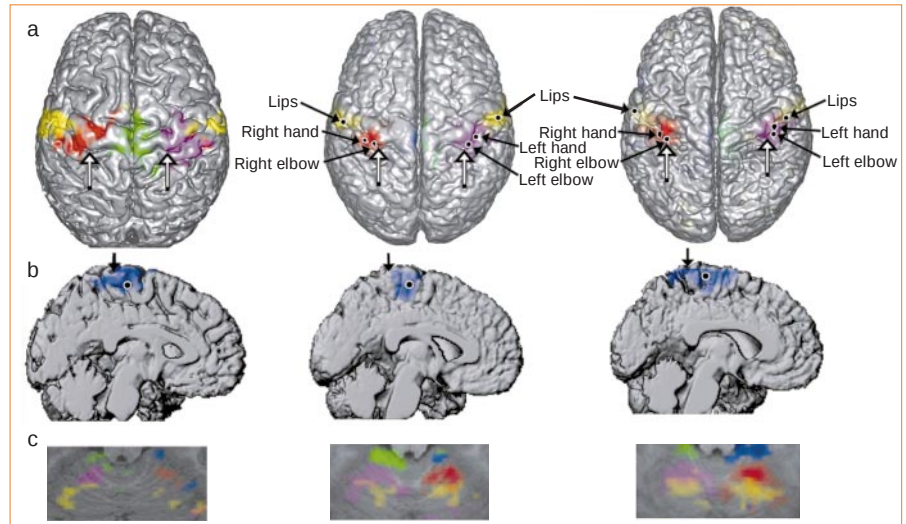


Figure 1 Cortical and cerebellar activation associated with attempted and executed movements in a paralysed individual (left), paralysed group (middle) and control group (right). **a, b**, Scaled integrals of t -test statistics (1.2 cm beneath the cortical surface); black dots indicate the location of t -value maxima. In **a** (top view), blue areas of activation are omitted from the left-side panel to avoid overlap; hollow arrows, putative hand-notches on the precentral gyri; **b**, mid-sagittal renderings of left-hemisphere activation resulting from attempted flexion of right toes. Arrows, central sulci. **c**, Activation maps (SPM(t)) in axial sections through the anterior cerebellum (21 mm inferior to the anterior–posterior commissural line). Red, abduction–adduction of fingers on the right hand, and purple, left hand; blue, flexion of right toes, and green, left toes; yellow, lip pursing.

motor and parietal regions). The attempted toe flexion and knee extension (five subjects) resulted in bilateral activation (stronger on the contralateral side) along the medial aspect of the longitudinal fissure (extending along the entire paracentral lobule). Executed movements of the elbows (results not shown) gave activation patterns that overlapped the one observed during attempted hand movements (elbow activation patterns were more dorso-posterior). Lip pursing resulted in activation on the most lateral aspect of the sensorimotor cortex.

Further consistent, organized activation occurred in the anterior cerebellum (Fig. 1c), the dorsal cingulate motor area, and an area extending on the banks of the sylvian fissure from the central sulcus (SII) to the operculum and the insula (results not shown). All of these activation features were reproduced in the control group (Fig. 1b, right panel) and are consistent with previous studies of cortical¹² and cerebellar¹³ motor maps. Although a consistent dorso-posterior shift of hand and elbow activation maxima was evident between injured and control groups, a similar shift was seen in the location of the group hand-knob, indicating that it may be due to intergroup anatomical differences.

We have shown that volitional activation of cortical sensorimotor representations in sufferers of paralysis occurs with only minimal reorganization of the gross somatotopy in subjects within five years of spinal-cord

injury. The remnant motor representations in these subjects seem to respond to intent, one of the principal physiological requirements for the development of a brain–computer interface device. Rapid progress in neural-interface technology may soon enable implantable neuroprosthetics to provide real-time reconstruction of complex voluntary motor activity for the paralysed.

Shy Shoham*, Eric Halgren†, Edwin M. Maynard*, Richard A. Normann*

**Department of Bioengineering, University of Utah, Salt Lake City, Utah 84112, USA*

e-mail: normann@utah.edu

†Nuclear Magnetic Resonance Center, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA

- Kennedy, P. R., Bakay, R. A. E., Moore, M. M., Adams, K. & Goldwithe, J. *IEEE Trans. Rehab. Eng.* **8**, 198–202 (2000).
- Wessberg, J. et al. *Nature* **408**, 361–365 (2000).
- Nicolelis, M. A. L. *Nature* **409**, 403–407 (2001).
- Green, J. B., Sora, E., Bialy, Y., Ricamato, A. & Thatcher, R. W. *Neurology* **53**, 736–743 (1999).
- Lotze, M., Laubis-Hermann, U., Topka, H., Erb, M. & Grodd, W. *Restor. Neurol. Neurosci.* **14**, 183–187 (1999).
- Sanes, J. N. & Donoghue, J. P. *Annu. Rev. Neurosci.* **23**, 393–415 (2000).
- Raineteau, O. & Schwab, M. E. *Nature Rev. Neurosci.* **2**, 263–273 (2001).
- Cheney, P. D. et al. *Prog. Brain Res.* **128**, 135–160 (2000).
- Turner, J. A. et al. *IEEE Trans. Neural Syst. Rehab. Eng.* **9**, 154–160 (2001).
- Friston, K. J. et al. *Hum. Brain Mapping* **2**, 189–210 (1995).
- Yousry, T. A. et al. *Brain* **120**, 141–157 (1997).
- Fink, G. R., Frackowiak, R. S., Pietrzyk, U. & Passingham, R. E. *J. Neurophysiol.* **77**, 2164–2174 (1997).
- Nitschke, M. F., Kleinschmidt, A., Wessel, K. & Frahm, J. *Brain* **119**, 1023–1029 (1996).

Ornithology

Arctic waders are not capital breeders

Birds prepare their eggs from recently ingested nutrients ('income' breeders) or from body stores ('capital' breeders)¹. As summers are short at Arctic latitudes, Arctic migrants have been presumed to bring nutrients for egg production from their previous habitats, so that they can start breeding immediately upon arrival¹⁻³. But we show here that eggs laid by 10 different wader species from 12 localities in northeast Greenland and Arctic Canada are produced from nutrients originating from tundra habitats, as inferred from carbon stable-isotope ratios in eggs, natal down, and juvenile and adult feathers.

During winter and migration, most Arctic-breeding waders eat estuarine invertebrates, shifting to terrestrial and limnic invertebrates on tundra breeding grounds. Invertebrates from estuarine and tundra habitats have distinctly different carbon-isotope ratios⁴. Diet-based differences in carbon-isotope ratios are expressed in bird tissues, including eggs and feathers^{4,5}. We investigated whether Arctic-breeding wader species that use estuarine habitats during the non-breeding season are capital breeders, producing eggs and hatchlings with carbon-isotope ratios that are typical of estuaries.

As expected, the flight feathers produced during early winter show ratios that

are typical of estuarine systems (Fig. 1). The same is true for shoulder feathers, which are among the last to be moulted before migration to breeding grounds⁶. However, the carbon-isotope signatures of eggs and natal down are typical of terrestrial and limnic systems. The eggs and resultant hatchlings therefore seem to be produced from local nutrients. This conclusion is supported by the strong resemblance of the carbon-isotope signatures of natal down and juvenile flight feathers, showing that eggs and post-hatch tissues have the same nutritional source.

Our results also indicate that the use of a mixed capital/income breeding strategy by these waders is unlikely. Egg-laying females use a single, local nutrient source, as indicated by the low within-clutch variation compared with the among-clutch variation in carbon-isotope ratios of natal down. (We used a hierarchically nested design in a mixed-model ANOVA, in which brood is nested within species as a random factor, and found significant effects of species ($F_{6,30} = 4.8$, $P < 0.002$) and brood ($F_{29,93} = 16.8$, $P < 0.001$) on the carbon stable-isotope ratios of natal down.)

We conclude that, with respect to egg production, the Arctic-breeding waders investigated here are income, rather than capital, breeders. For these waders, the fitness costs of transporting extra nutrient stores to breeding grounds⁷ outweigh the potential benefits¹⁻³. The capital strategy may still be used by large species of Arctic breeding migrants, such as geese^{1,2} — it is likely that larger species need relatively

smaller body stores for egg production. Also, larger species, although constrained by the same fixed period of opportunity as smaller birds, require longer to complete their breeding.

Marcel Klaassen*, Åke Lindström†‡, Hans Møller§, Theunis Piersma‡

*Centre for Limnology, Netherlands Institute of Ecology, PO Box 1299, 3600 BG Maarssen, The Netherlands

e-mail: klaassen@cl.nioo.knaw.nl

†Department of Animal Ecology, Ecology Building, Lund University, 22362 Lund, Sweden

‡Netherlands Institute for Sea Research, PO Box 59, 1790 AB Den Burg, Texel, and Centre for Ecological and Evolutionary Studies, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands

§Department of Arctic Environment, National Environmental Research Institute, PO Box 358, 4000 Roskilde, Denmark

1. Drent, R. H. & Daan, S. *Ardea* **68**, 225–252 (1980).
2. Newton, I. in *Evolutionary Ecology* (eds Stonehouse, B. & Perrins, C.) 113–126 (Macmillan, London, 1977).
3. Davidson, N. C. & Evans, P. R. in *Acta XIX Congressus Internationalis Ornithologici* (ed. Ouellet, H.) 342–352 (Univ. Ottawa Press, Ottawa, 1988).
4. Peterson, B. J. & Fry, B. *Annu. Rev. Ecol. Syst.* **18**, 293–320 (1987).
5. Hobson, K. A. *Oecologia* **120**, 314–326 (1999).
6. Cramp, S. (ed.) *Handbook of the Birds of Europe, the Middle East and North Africa* Vol. 3 (Oxford Univ. Press, Oxford, 1983).
7. Witter, M. S. & Cuthill, I. C. *Phil. Trans. R. Soc. Lond. B* **340**, 73–92 (1993).
8. Stettenheim, P. in *Avian Biology* Vol. 2 (eds Farner, D. S., King, J. R. & Parkes, K. C.) 1–63 (Academic, New York, 1972).

Intracellular signalling

Key enzyme in leptin-induced anorexia

Leptin is a key hormonal regulator of energy balance that acts upon hypothalamic neurons to reduce food intake, but the intracellular mechanisms involved are incompletely understood. Here we show that systemic administration of leptin in rats activates the enzyme phosphatidylinositol-3-OH kinase (PI(3)K) in the hypothalamus and that intracerebroventricular (i.c.v.) infusion of inhibitors of this enzyme prevents leptin-induced anorexia. Our results indicate that PI(3)K is a crucial enzyme in the signal-transduction pathway that links hypothalamic leptin to reduced food intake.

Leptin, a hormone derived from adipocytes (fat cells), is known to activate PI(3)K in non-neuronal cells and to elicit cellular responses that are blocked by PI(3)K inhibitors *in vitro*^{1,2}; it also activates the transcription factor STAT3 (ref. 3). To determine whether leptin activates PI(3)K, as well as STAT3, in the hypothalamus *in vivo*, we observed the effect of systemically administered leptin on their activity in male Wistar rats. As before³, leptin induced rapid tyrosine-phosphorylation of STAT3

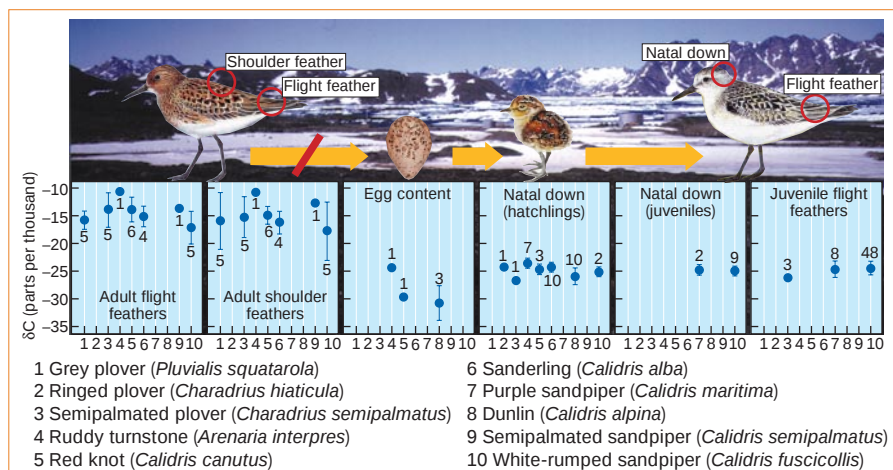


Figure 1 Carbon stable-isotope ratios (parts per thousand difference from $^{13}\text{C}/^{12}\text{C}$ ratio in Pee Dee limestone⁴; δC) of eggs, natal down and feathers of different species of Arctic-breeding waders at different times during the year. All samples were collected in northeast Greenland and Arctic Canada in 1999 and 2000. Feathers collected from nest-attending adults were grown either in winter (adult flight feathers) or during spring migration (adult shoulder feathers); data are averages across individuals $\pm$ s.d. Eggs were collected from deserted nests (egg content); natal down was collected from hatchlings; data are averages across clutches $\pm$ s.d. per species. Data for natal down still attached to the tips of head and neck feathers of independent young⁹, and for secondary flight feathers from independent young are averages across individuals $\pm$ s.d.). Values of δC were determined using a Carlo Erba model 1106 elemental analyser which was coupled online, through a Finnigan conflow 2 interface, to a Finnigan Delta S mass spectrometer. Feathers were washed in chloroform before carbon-isotope analysis. Using a hierarchically nested design in a mixed model ANOVA, in which species was nested within tissue type, which in turn was nested within the tissue groups 'adult feathers' and 'egg, natal down and juvenile feathers', we found significant effects of tissue group ($F_{1,152} = 805.3$, $P < 0.001$), tissue type ($F_{4,152} = 6.7$, $P < 0.001$) and species ($F_{9,152} = 2.8$, $P < 0.005$) on δC .

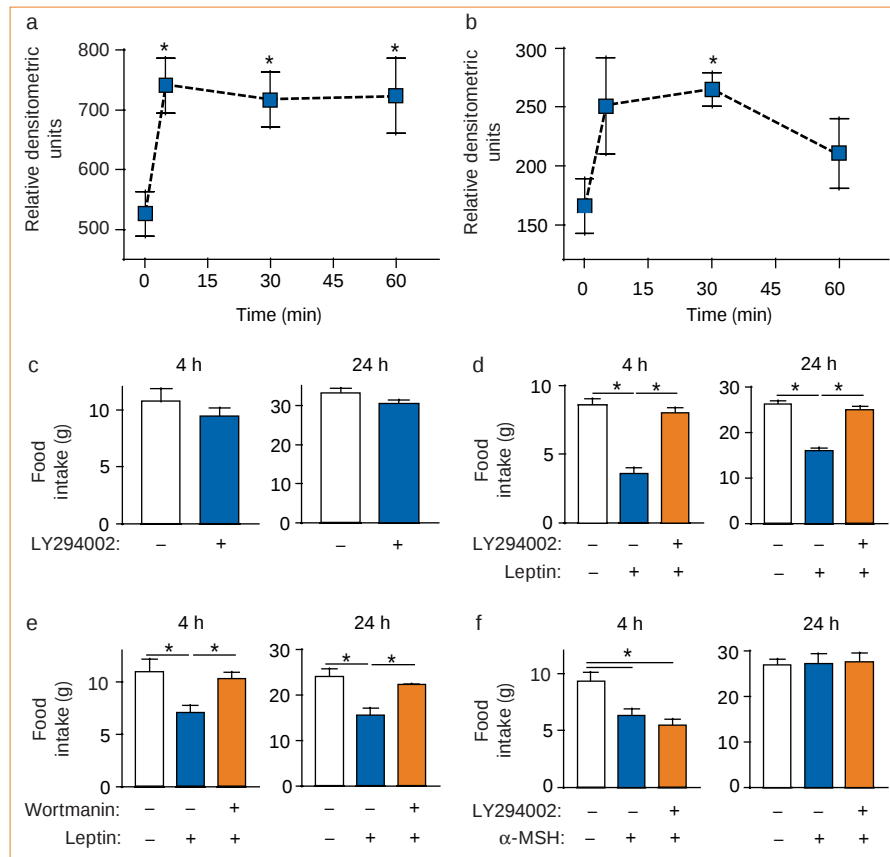


Figure 1 Role of PI(3)K signalling in the anorectic response to leptin. Time course of **a**, tyrosine-phosphorylation of STAT3, and **b**, IRS-2-associated PI(3)K activity¹² in mediobasal hypothalamic extracts after intraperitoneal injection of leptin (1 µg g⁻¹; provided by A. F. Parlow) into male Wistar rats. **c**, LY294002 (1 nmol) injected intracerebroventricularly (i.c.v.) at the onset of the dark cycle does not alter food intake at 4 h or 24 h. **d-f**, Pretreatment with LY294002 (1 nmol i.c.v.; **d**, **f**), wortmannin (0.01 nmol; **e**) or vehicle (**d-f**) was followed 1 h later, just before the onset of the dark cycle, by i.c.v. administration of leptin (3 µg; **d**, **e**), α-MSH (25 µg; **f**) or vehicle (**d-f**). Pretreatment with i.c.v. LY294002 (**d**) or wortmannin (**e**) blocks leptin-induced anorexia, whereas LY294002 does not affect α-MSH-induced anorexia (**f**). *n* = 6–8 rats per group; asterisk denotes *P* < 0.05 relative to controls (Student's *t*-test or one-way ANOVA).

in mediobasal hypothalamic extracts, an effect that was sustained for up to 1 h (Fig. 1a). This response was paralleled by a comparable increase in hypothalamic PI(3)K activity associated with the insulin-receptor substrate IRS-2, peaking at 30 min (Fig. 1b) (IRS proteins couple activated cell-surface receptors of the tyrosine-kinase type to PI(3)K and other signalling molecules).

Pharmacological inhibitors of PI(3)K activity, such as LY294002 and wortmannin, can block leptin activity *in vitro*^{1,2}, suggesting that PI(3)K-dependent signalling is needed for certain intracellular responses to leptin. We tested whether the inhibitory effects of leptin on food intake depend on PI(3)K, by i.c.v. infusion of PI(3)K inhibitors into rats. We first identified a dose of LY294002 (1.0 nanomol i.c.v.; results not shown) that did not alter food intake when administered to male Wistar rats at the onset of the dark cycle (Fig. 1c) — the period of the day during which rats consume the most food — and then evaluated the effect of i.c.v. pretreatment with this dose of inhibitor or its vehicle on the anorectic response to i.c.v. leptin.

Treatment with leptin markedly reduced

food intake at both 4 h and 24 h in rats pretreated with vehicle, but we did not detect leptin-induced anorexia after pretreatment with LY294002 (Fig. 1d). By using wortmannin instead of LY294002 in a different group of rats, we confirmed that infusion of PI(3)K inhibitors into the third cerebral ventricle prevents the reduction in food intake induced by i.c.v. administration of leptin (Fig. 1e). As PI(3)K is the only enzyme that is known to be inhibited by both LY294002 and wortmannin⁴, these findings indicate that neuronal PI(3)K is important for the effects of leptin on food intake.

Melanocortin peptides such as melanocyte-stimulating hormone (α-MSH) also reduce food intake, but act through a different pathway from that used by leptin, involving intracellular cyclic AMP and protein kinase A⁵. If the effect of PI(3)K inhibitors on leptin-induced anorexia depends only on PI(3)K, then they should not affect the anorectic response to melanocortins. We therefore tested the effect on animals pretreated with i.c.v. LY294002 of injecting i.c.v. α-MSH and found that food intake was still reduced 4 h later (Fig. 1f). As the inhibitory effect of α-MSH

on food consumption is short-lived⁶, there was no difference between groups after 24 h.

These findings show that leptin activates hypothalamic PI(3)K signalling *in vivo* and that leptin, but not melanocortins, reduces food intake through a PI(3)K-sensitive mechanism. Our results are consistent with a model in which binding of leptin to its receptor on hypothalamic neurons stimulates intracellular signalling through PI(3)K, as occurs in non-neuronal cells^{1,2}, although leptin-induced activation of hypothalamic PI(3)K may also occur in neurons further 'downstream' in a leptin-activated pathway.

Like leptin, the pancreatic hormone insulin may function as an afferent signal to the hypothalamus in the regulation of body-fat levels⁷. As PI(3)K is an intracellular mediator of insulin activity in peripheral tissues⁸ and hypothalamic neurons⁹, our results raise the possibility that insulin and leptin can both activate neuronal IRS-PI(3)K signalling. This would not only explain the overlapping hypothalamic activities of these two hormones^{7,9}, but may also have implications for the development of hypothalamic leptin resistance, which is associated with obesity.

Insulin-stimulated activation of PI(3)K is impaired in the peripheral tissues of obese individuals, which may contribute to the pathogenesis of obesity-induced insulin resistance⁸. If the mechanism used by leptin to reduce food intake is PI(3)K-dependent, as our results suggest, defective activation of PI(3)K in hypothalamic neurons may reduce the ability of leptin to promote weight loss in obese rodents¹⁰ and humans¹¹.

Kevin D. Niswender*, Gregory J. Morton*, Walter H. Stearns†, Christopher J. Rhodes‡, Martin G. Myers Jr†, Michael W. Schwartz*

*Division of Metabolism, Endocrinology and Nutrition, University of Washington School of Medicine and Harborview Medical Center, Seattle, Washington 98104, USA

e-mail: mschwartz@u.washington.edu

†Joslin Diabetes Center, Harvard Medical School, Boston, Massachusetts 02215, USA

‡Pacific Northwest Research Institute, Seattle, Washington 98122, USA

- Harvey, J. et al. *J. Biol. Chem.* **275**, 4660–4669 (2000).
- Zhao, A. Z. et al. *J. Biol. Chem.* **275**, 11348–11354 (2000).
- Vaisse, C. et al. *Nature Genet.* **14**, 95–97 (1996).
- Davies, S. P., Reddy, H., Caivano, M. & Cohen, P. *Biochem. J.* **351**, 95–105 (2000).
- Cone, R. D. et al. *Recent Progr. Horm. Res.* **51**, 287–317 (1996).
- McMinn, J. E., Wilkinson, C. W., Havel, P. J., Woods, S. C. & Schwartz, M. W. *Am. J. Physiol. Reg. Integr. Comp. Physiol.* **279**, 695–703 (2000).
- Woods, S. C., Seeley, R. J., Porte, D. Jr & Schwartz, M. W. *Science* **280**, 1378–1383 (1998).
- Shulman, G. I. *J. Clin. Invest.* **106**, 171–176 (2000).
- Spanwick, D., Smith, M. A., Mirshamsi, S., Routh, V. H. & Ashford, M. L. *Nature Neurosci.* **3**, 757–758 (2000).
- El-Haschimi, K., Pierroz, D. D., Hileman, S. M., Bjorbaek, C. & Flier, J. S. *J. Clin. Invest.* **105**, 1827–1832 (2000).
- Heymsfield, S. B. et al. *J. Am. Med. Assoc.* **282**, 1568–1575 (1999).
- Myers, M. G. Jr et al. *Proc. Natl Acad. Sci. USA* **89**, 10350–10354 (1992).

Weakening link to colorectal cancer?

The catalytic γ -subunit of the enzyme phosphatidylinositol-3-OH kinase (PI(3)K γ) relays signals from G-protein-coupled receptors at the cell membrane and mediates leukocyte responses to chemokines and chemoattractants^{1–3}. Sasaki *et al.*⁴ report that mice that cannot produce PI(3)K γ have a high incidence of colorectal carcinomas, causing weight loss and premature death. However, PI(3)K γ -null mouse strains have been independently generated in three other laboratories; none of these mice developed tumours and their weight and lifespan were normal. This casts doubt on the idea that loss of functional PI(3)K γ leads directly to transformation of colon epithelial cells and tumour progression.

Disrupting signalling by chemokine receptors has been considered as a strategy to fight chronic inflammatory disease. Signals from these receptors are integrated by PI(3)K γ ^{5,6}, whose crystal structure⁷ and

inhibitor interactions⁸ are understood in detail, paving the way to rapid therapeutic exploitation of PI(3)K γ as a drug target. But promising research was interrupted by the claim of Sasaki *et al.*⁴ that loss of functional PI(3)K γ causes colon cancer in mice.

Sasaki *et al.*⁴ base their conclusions on the fact that their PI(3)K γ -null mouse strain rapidly developed colorectal carcinomas. Using total colon and mucosal samples, they detected PI(3)K γ in colon tissue but not in murine or human colorectal adenocarcinomas, inferring that the loss of PI(3)K γ was crucial to the transformation process in epithelial cells.

The murine PI(3)K γ gene was independently inactivated by four groups, including ourselves^{1–3} and B.L. *et al.* (unpublished observations), using four different strategies (Fig. 1a), all of which confirmed that this enzyme is important for transmission of inflammatory signals. In our studies, however, mice lacking PI(3)K γ did not develop tumours, or succumb to weight loss and premature death (Fig. 1b). Analysis of tissue biopsies from more than 100 PI(3)K γ -null mice at various ages and of both sexes from two genetic backgrounds (129/Sv inbred and

C57BL/6J/129 outbred) showed no malignant transformation (results not shown).

We therefore re-examined the PI(3)K γ -expression pattern reported by Sasaki *et al.*⁴, and found that PI(3)K γ signals in colonic mucosa correlate with the presence of leukocytes, as shown by the CD18 marker or by histology, but that PI(3)K γ is undetectable in normal colonic epithelial cells (positive for the Lu5 cytokeratin marker) from mice, human patients or rats (Fig. 1c–e). We conclude that normal and transformed colonic epithelial cells (such as the HT29 cancer cell line) do not express detectable amounts of PI(3)K γ , making a direct cause-and-effect relationship between loss of PI(3)K γ and development of colon cancer unlikely.

Invasiveness and growth-factor-independent survival of human colorectal HCT8/S11 tumour cells was promoted by constitutively active, membrane-targeted PI(3)K γ (PI(3)K γ -CAAX), but not by its absence or by stable transfection with catalytically inactive PI(3)K γ (KR-CAAX; Fig. 1f). This suggests that malignancy is coupled to activated PI(3)K and not to its loss.

Our findings are consistent with a lack of tumorigenesis in PI(3)K γ -null strains generated by three out of four strategies. The reproducibility and consistency of the diverging results, however, make it possible that an unknown gene-targeting effect enhances other growth-promoting signals in the PI(3)K γ -null allele used by Sasaki *et al.*⁴. Their interesting phenotype therefore needs further investigation, and may eventually reveal an important cause of colon cancer.

M. Barbier*, S. Attout†, R. Calvez*, M. Laffargue*, A. Jarry‡, M. Marell§, F. Altruda||, C. Gespach†, D. Wu¶, B. Lu#, E. Hirsch||, M. P. Wymann*

*Institute of Biochemistry, University of Fribourg, 1700 Fribourg, Switzerland

e-mail: matthiaspaul.Wymann@UniFR.ch

†INSERM U482, Hôpital Saint-Antoine,

75571 Paris 12, France

‡INSERM U539, Hôtel-Dieu, 44035 Nantes, France

§Laboratory of Experimental Cancerology, Ghent

University, De Pintelaan 185, 9000 Ghent, Belgium

||Dipartimento di Genetica, Biologia e Biochimica,

Università di Torino, 10126 Torino, Italy

¶Department of Genetics and Developmental

Biology, University of Connecticut, Farmington,

Connecticut 06030, USA

#Children's Hospital, Harvard Medical School,

Boston, Massachusetts 02115, USA

1. Hirsch, E. *et al.* *Science* **287**, 1049–1053 (2000).
2. Sasaki, T. *et al.* *Science* **287**, 1040–1046 (2000).
3. Li, Z. *et al.* *Science* **287**, 1046–1049 (2000).
4. Sasaki, T. *et al.* *Nature* **406**, 897–902 (2000).
5. Wymann, M. P., Sozzani, S., Altruda, F., Mantovani, A. & Hirsch, E. *Immunol. Today* **21**, 260–264 (2000).
6. Rickert, P., Weiner, O. D., Wang, F., Bourne, H. R. & Servant, G. *Trends Cell Biol.* **10**, 466–473 (2000).
7. Walker, E. H., Perisic, O., Ried, C., Stephens, L. & Williams, R. L. *Nature* **402**, 313–320 (1999).
8. Walker, H. E. *et al.* *Mol. Cell* **6**, 909–919 (2000).
9. Hirsch, E. *et al.* *Gene* **256**, 69–81 (2000).
10. Branka, J. E. *et al.* *Gastroenterology* **112**, 1887–1894 (1997).
11. Bondeva, T. *et al.* *Science* **282**, 293–296 (1998).

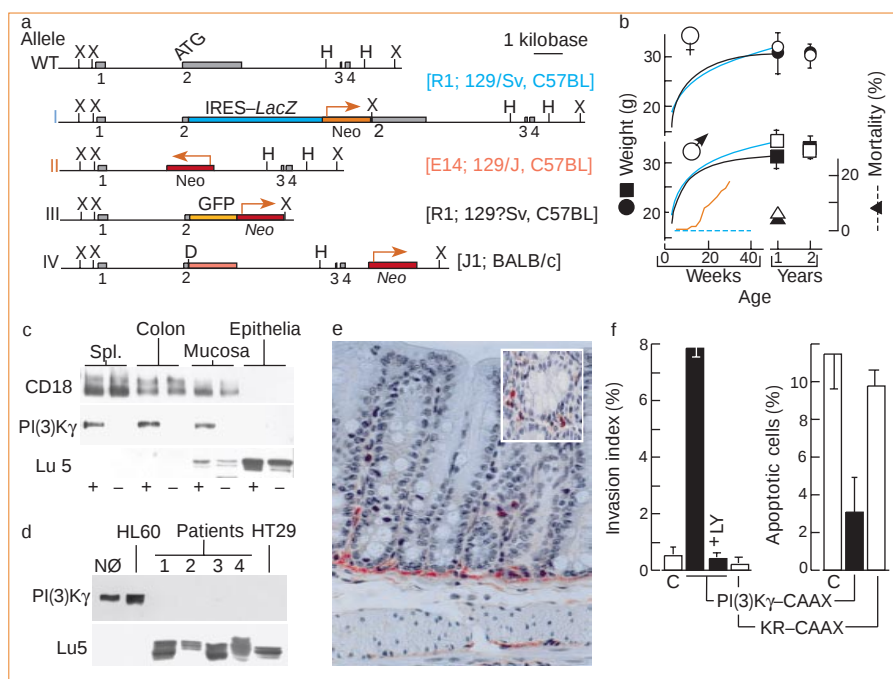


Figure 1 Gene targeting and expression of phosphatidylinositol-3-OH kinase γ -subunit (PI(3)K γ). **a**, The wild-type PI(3)K γ allele (WT) was targeted using four strategies. Allele I: interception¹ of exon 2 (embryonic stem cells and mouse strains indicated; for details of the PI(3)K γ gene, see ref. 9); allele II: deletion of exon 2, as carried out by Sasaki *et al.*⁴; direction of transcription of Neo is opposite to that of PI(3)K γ ; allele III: fusion³ of PI(3)K γ to green fluorescent protein (GFP) and excision of exon 2; allele IV: start of exon 2 deleted (~350 base pairs; B.L. *et al.*, unpublished observations). **b**, Weight and mortality of normal (filled symbols; black lines) and PI(3)K γ ^{-/-} (allele I: white symbols; blue lines) mice; 465 WT and 602 PI(3)K γ ^{-/-} animals; age, 2 yr; sex ratio, 9/10 WT and 34/8 PI(3)K γ ^{-/-} females/males; mortality from Sasaki *et al.*⁴ shown in red. **c**, Murine WT (+) and PI(3)K γ ^{-/-} (-) splenocytes (spl.), total colon, mechanically sheared mucosa and isolated colon epithelial cells were probed for the leukocyte marker CD18, PI(3)K γ (mouse monoclonal anti-PI(3)K γ , amino-terminal epitope), and the pan-epithelial marker Lu-5. **d**, Anti-PI(3)K γ and Lu-5 western blots of total human cell lysates from neutrophils (N2), retinoic-acid-differentiated HL60, primary cultures of normal human colonocytes¹⁰ from large bowel resected from four patients with diverticulitis, and the HT29-CI.16E cell line. **e**, PI(3)K γ immunoreactivity (red) in rat colonic mucosa (inset, cross-section of crypt). **f**, Invasion of collagen gels and serum-withdrawal-induced apoptosis of human colorectal HCT8/S11 cells stably transfected with control vector (C), membrane-targeted PI(3)K γ (PI(3)K γ -CAAX, black bars¹¹) or a catalytically inactive mutant (PI(3)K γ (K833R)-CAAX; KR-CAAX, white bars). The PI(3)K inhibitor LY294002 (LY) was used at 10 μ M. Further details are available at <http://www.unifr.ch/biochem/wymann>.

Morphogen gradient interpretation

J. B. Gurdon & P.-Y. Bourillot

Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, UK, and Department of Zoology, University of Cambridge, Cambridge, UK

A morphogen gradient is an important concept in developmental biology, because it describes a mechanism by which the emission of a signal from one part of an embryo can determine the location, differentiation and fate of many surrounding cells. The value of this idea has been clear for over half a century, but only recently have experimental systems and methods of analysis progressed to the point where we begin to understand how a cell can sense and respond to tiny changes in minute concentrations of extracellular signalling factors.

Morphogens are secreted signalling molecules that organize a field of surrounding cells into patterns. They form a gradient of concentration emanating from a localized source, and determine the arrangement and fate of responding cells according to the different concentrations of the morphogen perceived by the cells. The idea of a morphogen gradient is intimately associated with the concept of positional information¹. A cell is believed to read its position in a concentration gradient of an extracellular signal factor, and to determine its developmental fate accordingly^{2,3}.

Morphogen action is of special importance in understanding development. This is because a single event, the emission of morphogen from a source, can lead to the formation of several different cell types in a correct spatial relationship to each other. This is a highly efficient way of creating complex patterns of gene expression and spatial position from a population of uncommitted cells in an embryo.

An understanding of morphogen gradients requires answers to two different questions. The first asks how a desired concentration gradient is formed. We need to know the identity of the morphogen, the shape and absolute concentration of the gradient, and the factors that create and maintain the gradient in its various positions. Much current work is providing some answers to these questions, especially by identifying extracellular molecules that bind morphogens, and that thereby influence the concentration of morphogen free to reach cells whose responses they determine. This subject has been widely reviewed^{4,5}, and we summarize here, only briefly, some of the principles that are emerging.

The second question asks how cells interpret a morphogen concentration. To understand this, we need to know how cells recognize different threshold concentrations of morphogen through receptors on their surface and how they transduce this information to the nucleus to create the appropriate gene or cell fate response. This second question has so far been little explored, and is the one we primarily address here.

Examples of morphogens

To be sure that an example of morphogen action exists, several criteria need to be satisfied. Ideally an identified signalling molecule should be present in the right place at the right time. It should be shown to be released from a localized source, forming a concentration gradient over a population of nearby and distant cells that respond directly to the signalling molecule in a concentration-dependent way. Cells in the pathway of the gradient should show two or more qualitatively different responses, such as the expression of different genes, in addition to their default pathway. Over- and underexpression experiments should have the effect of changing gene expression or cell fate in the predicted directions. Thus an increase in the gradient, such that all cells experience an elevated morphogen concentration, should cause them to switch their response to a higher level. Likewise, underexpression should cause

cells to move progressively down the scale of responses available to them. A further important characteristic of morphogen action is that it should be direct. This means that wherever a cell is located in the concentration gradient, it should respond directly to the same morphogen, and should not judge its position indirectly through the mediation of other signal molecules or other cells elsewhere in the responding population.

The most convincing and best-analysed examples of morphogen action are listed in Table 1. Examples of morphogens acting during development in *Drosophila* and *Xenopus* are illustrated in Figs 1 and 2. In all of these cases, the morphogens are secreted proteins that are functional only when cut, and often further modified, after synthesis of the primary polypeptide from a gene. Where the most profound analysis has been conducted, it is usually found that a primary morphogen gradient is supplemented, or even created by, other positively or negatively acting factors, such as Short gastrulation (Sog) in the *Drosophila* wing⁶ or chordin in the amphibian embryo⁷ (see Table 1). These agents affect the concentrations of the factor (in these cases, Decapentaplegic (Dpp) or bone morphogenetic protein (BMP), respectively) that reach the surface of responding cells, and to which cells respond directly when interpreting their position in the gradient. The table does not include Bicoid (*Drosophila* embryo), which has many properties of a morphogen⁸, except that it is a transcription factor that can spread through the syncytial embryo, a situation different from other morphogens that are extracellular and spread through multicellular tissue. We concentrate here on the *Drosophila* wing disc and the amphibian blastula animal cap, two systems in which the principles of morphogen gradient interpretation have been extensively studied.

Most of the examples in Table 1 fall short, in one or more respects, of the perfect morphogen. Indeed, few if any of the morphogens so far described obey all the rules laid down by theoreticians for their guidance. Only in a few cases has a natural morphogen gradient been directly visualized (see, for example, refs 9–11); in other cases, the existence of a gradient has to be deduced from its activity, that is, its ability to induce genes in a concentration-related sequence. This difficulty arises because proteins that function as morphogens seem to have activity at extremely low concentrations (for example, activin at 50 pM), and because extracellular diffusible proteins are hard to fix for *in situ* demonstration. For some examples, a read-out of gradient interpretation is satisfactorily provided by directly induced gene expression (such as the *Drosophila* wing and *Xenopus* embryo (Figs 1 and 2)). In others (such as the chick limb) there are no position-related gene or cell-type markers, and gradient effects have to be judged from overall organ or tissue morphology (for example digit number). Here, we concentrate on the former. Other examples of morphogens may exist, such as retinoic acid in the vertebrate anteroposterior neural axis¹², but in most cases they have not been shown to have direct action.

The strongest candidate morphogens are members of the

transforming growth factor- β (TGF- β), Hedgehog (Hh) and Wnt families of secreted proteins. Members of the fibroblast growth factor (FGF) family seem to act as competence factors having permissive rather than concentration-related effects¹³.

Gradient formation

Much recent work has been directed towards the problem of how the necessary amount of morphogen comes to be located in the right place at the right time (see ref. 5 for original work in this field). A difficulty is that we are poorly informed about the actual concentration and shape of the gradient itself. Known morphogens seem to be effective at extremely low concentrations, 10^{-9} to 10^{-11} M, and are probably not distributed in an even slope across their field of action. Only in a few cases can a glimpse of the morphogen itself be caught by the use of tagged forms containing green fluorescent protein (GFP) or horseradish peroxidase, or by antibody staining. In most cases the source of the morphogen is evident from the distribution of its messenger RNA. But we need to know the rate at which the protein is released from the source, the rate of its spread and its stability. We also need to know whether the gradient is formed with the morphogen in its active state, or whether it is converted into an active condition during or after gradient formation. The complexity of the problem is well exemplified by Hh, which undergoes an endonuclease cleavage and needs cholesterol addition for full activity¹⁴. Particularly important is the recent finding that, more often than not, the activity of a morphogen is controlled by antagonistic proteins that associate with it and prevent it from binding to its receptors (Table 1). A further level of complexity seems to exist in controlling the distribution and activity of these antagonistic factors. For example, the activity of the BMP gradient in vertebrate embryos is determined primarily by an opposing gradient of chordin and other molecules, the activity of chordin being itself determined by the protease xoloid⁷. In the *Drosophila* wing disc, the different shape of the Wingless (Wg) gradient in the anterior and posterior regions is caused by faster endocytosis and degradation of Wg in posterior cells¹¹. It seems that the major factors shaping a gradient are not only different for each morphogen, but may also differ for the same morphogen in different stages of development. Other characteristics of gradients vary greatly from one case to another. For example, the range of a gradient, or the distance over which it has activity, can be as small as

a few cell diameters (50 μ m for Hh in the wing disc) or as large as 300 μ m (for the mouse limb-bud and amphibian embryo). To understand how cells read their position in a gradient, it is important to know its slope. Judging from responses of genes to gradients of Dpp and activin, we estimate that cells read a threefold concentration change over three cell diameters (about 30 μ m); however, the involvement of antagonistic factors may make the effective gradient much steeper than this.

There has been much activity in analysing the mechanism of transmission of a morphogen across its field⁵. Three ideas prevail: (1) diffusion in the extracellular matrix (active or facilitated); (2) relay by sequential internalization and re-emission from cell to cell; and (3) cytoplasmic contact by threads of cytoplasm connecting distant cells (cytonemes). The use of transplanted tissues defective for receptors or ligands has shown that TGF- β morphogens such as squint (fish)¹⁵ and activin (frogs)¹⁶ can spread by diffusion over long distances. In contrast, cells defective for endocytosis interfere with the formation of Dpp gradients, a result in support of a relay mechanism. It even seems that both of these mechanisms can have a role in forming the same gradient of Dpp in the wing disc.

Lastly we should point out that the timing of gradient formation is likely to be rapid. In later development, as in the *Drosophila* wing disc, a Dpp gradient is normally formed slowly, extending over 25 cell diameters in 3 days. Nevertheless, the same Dpp gradient can be reformed, after temperature interruption, at the rate of 4 cells in 1 h^{9,10}. In *Xenopus* embryos, however, an activin gradient can be formed experimentally over 100 μ m in 1 h¹⁷, and natural gradients are normally formed in *Xenopus* and *Drosophila* embryos in 2 h or less. We discuss below how cells interpret such rapid changes in morphogen concentration.

Perception at the cellular level

A simple mechanism by which a population of cells could make different responses to morphogen concentration would exist if cells have different response thresholds. According to this idea, each cell would have only a binary choice: respond to the morphogen or not. For each cell the actual concentration to which it would respond would be different. In this way, a range of concentration-related responses could be made by a cell population, even though any one cell would be able only to respond or not. This would avoid the

Table 1 Examples of morphogens

Developmental process	Signal source	Morphogen (range)*	Anti-factor	Receptors	Gene response† (concentration)	Gene response (time, h)	Reference
<i>Drosophila</i> embryo dorsoventral axis	Dorsal region	Scw (short) Dpp (long)	Sog Brinker	Tkv Sax Punt	<i>Race</i> (high) <i>zen</i> (middle) <i>pannier</i> (low)	3	19, 21, 47
<i>Drosophila</i> imaginal wing disc	Anteroposterior compartment boundary	Dpp (long)	Brinker	Tkv Punt	<i>sal</i> (high) <i>omb</i> (low)	24–72‡	6
<i>Drosophila</i> imaginal wing disc	Dorsoventral compartment boundary	Wg (long)		Fz	<i>neur</i> (high) <i>Dll</i> (middle) <i>vg</i> (low)	24–72‡	34, 50
<i>Xenopus</i> mesoderm formation	Nieuwkoop centre	Activin (long) Xnr1, -2, -6 (short)	Antivin Follistatin	Activin receptors I, II	<i>gsc</i> , <i>eomes</i> (high) <i>Xbra</i> , <i>apod</i> (low)	3	7
<i>Xenopus</i> axis formation	Organizer (Spemann)	BMP2, -4	Chordin Noggin Follistatin	BMP receptors	<i>Xvent</i> (high) <i>MyoD</i> (middle) <i>Xnot</i> (low)	3	7
Zebrafish axis formation	Organizer	Squint (long) Cyclops (short)	Lefty	Activin receptors I, II Oep (coreceptor)	<i>gsc</i> (high) <i>Ntl</i> (<i>Xbra</i>) (low)	3	15, 45
Vertebrate neurogenesis	Notochord; neural floor plate	Shh (long)	BMP4 and TGF- β family	Ptc Smoothed	<i>Nkx2.2</i> (high) <i>Dbx</i> (middle) <i>Pax7</i> (low)	12	42, 43

* Short range, 20 μ m or less; long range, 100 μ m or more.

† Includes repression as well as activation.

‡ Only after 50 h is *omb* expression further from the source than *sal*.

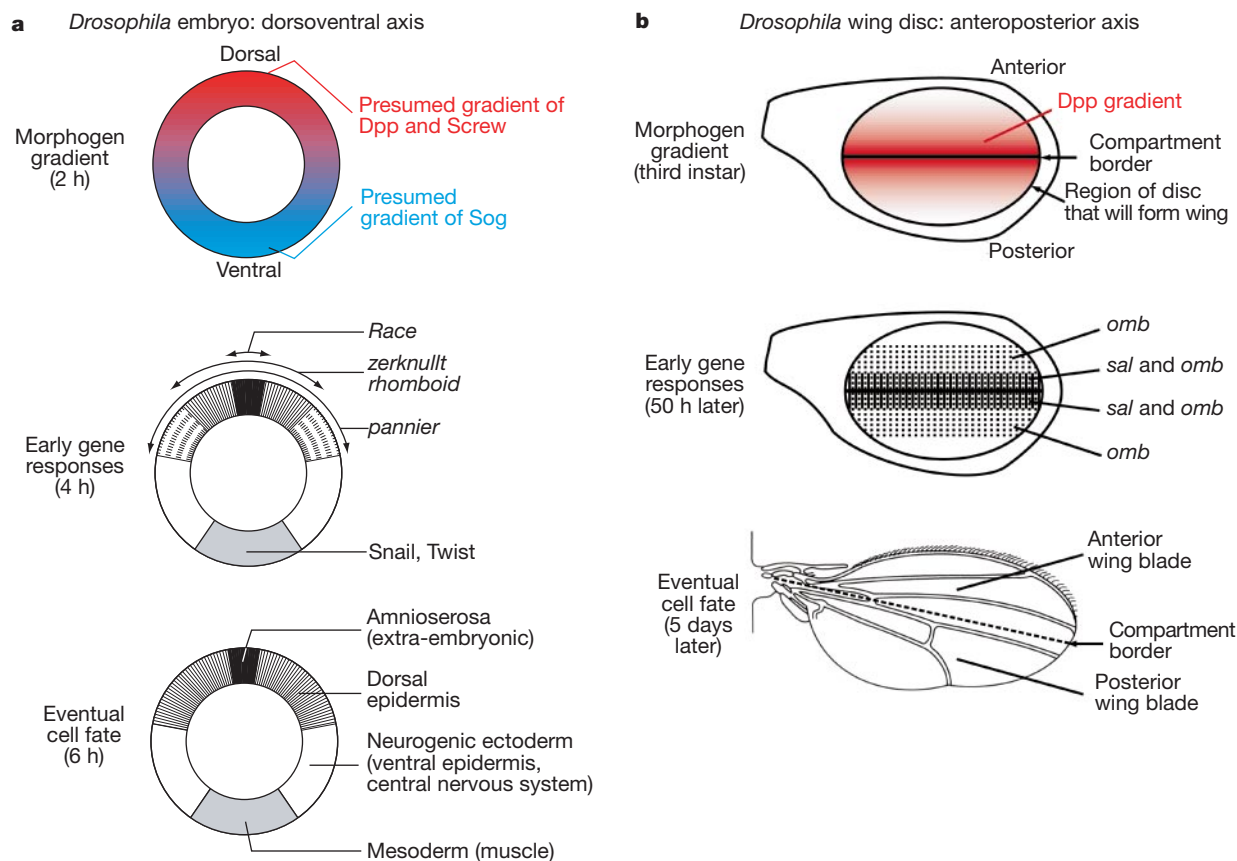


Figure 1 Morphogen action during development in *Drosophila*. **a**, Diagrams of cross-sections of *Drosophila* embryos, showing the presumed morphogen gradients, early gene responses and cell fates⁴⁸. Early gene responses overlap in the dorsal region. Gradients of Dpp and Screw, and their inhibitor Sog, are formed as an effect of a maternal ventral-to-dorsal gradient of the nuclear protein Dorsal. An intracellular gradient of the repressor Brinker coincides in position and cooperates functionally with the Sog gradient. **b**, Diagrams looking down on a *Drosophila* wing disc. The repressor Brinker is present in the wing-forming regions not occupied by Dpp. The transition from a third-instar disc to a complete wing involves complicated folding and extension movements.

need to credit cells with the ability to make any more sophisticated a response than an on/off switch.

In most cases it is not possible to be sure that any one cell has more than a binary choice of response. In a few examples, however, it seems clear that an individual cell chooses between at least three responses. One of these examples is when activin-loaded beads are placed in a spherical cell reaggregate of two face-to-face animal caps of a *Xenopus* blastula (Fig. 2b). There is no movement, and very little cell division, in the responding cell population; nevertheless a ripple of *Xbrachyury* (*Xbra*) gene expression spreads out radially from the activin beads, and this is followed 1–2 h later by the expression of *gooseoid* (*gsc*) and *omesodermin* (*omes*) in more centrally located cells. These gene responses are direct (cycloheximide insensitive) and expand radially as the concentration of activin on the beads, or as the time over which the beads are left in place, is increased¹⁷. In the same cells, *Xbra* expression is followed by *gsc* and *omes* expression, and subsequently by *sox* expression. *Xbra* suppression at higher activin concentrations is an indirect effect of *gsc* and other high-response genes¹⁸. The same choice of response seems to apply to the *Drosophila* embryo. A two- to fourfold change in Dpp by RNA injection is sufficient to cause the ectoderm cells, which do not change position relative to each other at this stage, to adopt one of three different fates^{19,20}. Over- or underexpression of Dpp causes the expression of *spalt* (*sal*) and *optomotor blind* (*omb*) genes to move nearer or further from the signal source in the wing disc²¹. Therefore, in all of these cases, the same cell has a choice of at least three

different responses to gradient concentration.

Another important question concerning the cellular basis of morphogen perception asks whether a cell needs its neighbours to determine its position in a gradient, or whether it can measure concentration on its own. In regeneration, such as occurs in newt limbs or cockroach legs, a missing structure is replaced by cells according to their position. Cells must therefore be guided by their neighbours to follow a particular fate, a concept commonly illustrated by Wolpert's French flag model²². The most decisive answer to this question requires a test in which isolated cells, unable to receive information from their neighbours, are exposed to different concentrations of morphogen, and tested for gene response. The only system in which this has been done is with *Xenopus* blastula cells exposed to activin, and then cultured either as single cells, confluent monolayers (lateral contact), or as reaggregates (contact on all sides)²³. These configurations have no effect on the type or amount of *Xbra* or *gsc* gene expression, and therefore show that, at least in this case, a cell can read its position in a concentration gradient without reference to its neighbours.

The fact that isolated cells have the ability to measure concentration independently of their neighbours does not necessarily mean that they do so in normal tissues. Almost certainly the spread of morphogen, and hence the exposure of a cell to the appropriate concentration of morphogen, is greatly affected by its neighbours. For example, underexpression of the Hh coreceptor Patched (Ptc) will increase the availability of Hh to nearby cells²⁴. However, the

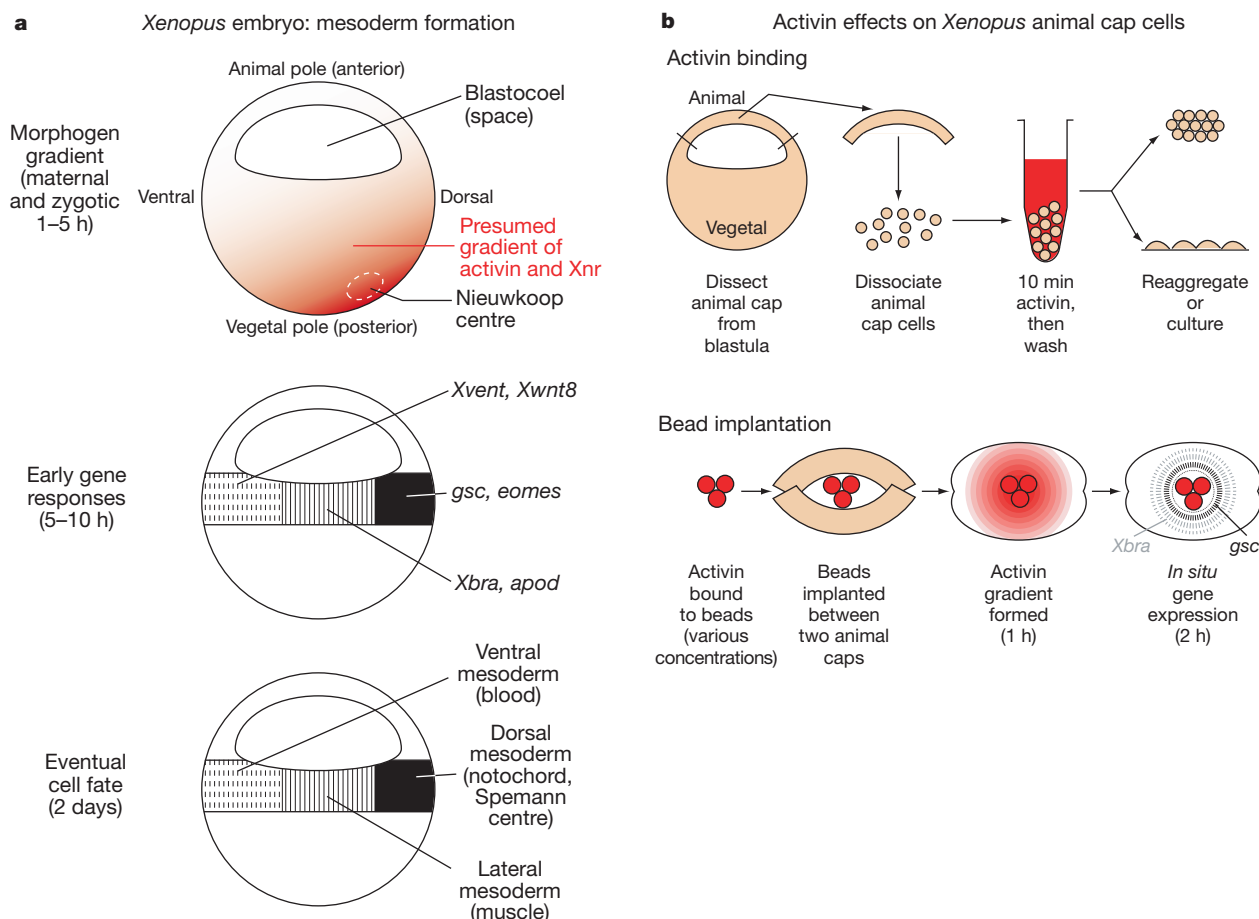


Figure 2 Examples of morphogen action during development in *Xenopus*. **a**, Diagrams of cross sections of an amphibian embryo. Gene responses are shown only for the equatorial mesodermal region, which is most directly responsive to Nieuwkoop signalling and to an activin or Xnr morphogen gradient. **b**, Assays used with amphibian animal cap cells to analyse interpretation of morphogen gradients. Activin is currently believed to simulate the action of endogenous *Xenopus* nodal proteins.

reading or interpretation of concentration by a cell seems to be independent of its neighbours. In *Drosophila* wing discs, clones of cells that overexpress the Dpp receptor Thickveins (Tkv) or other components of the Dpp pathway can be created in the Dpp gradient. In spite of this, cells express the induced genes *sal* and *omb* in accord with their genetic constitution and not with their position in the gradient²⁵. Therefore cells respond to Dpp signalling according to their own interpretation of gradient concentration, and not according to any influence from their neighbours. In *Xenopus*, activin can elicit a normal concentration response when passing through cells unable to respond to activin (endodermal cells) and through tissue inhibited for protein synthesis by cycloheximide¹⁷. These cells can not depend on their normal neighbours or engage in relay signalling in these cases. All these experiments argue that cells interpret position in a concentration gradient independently of their neighbours.

Similar transplantation or mosaic experiments have demonstrated, at least in some cases, the important point that morphogen action is direct. An example of indirect action is the release of Hh from a row of cells along the anteroposterior compartment border in *Drosophila* wing discs. Hh travels a very short distance and induces the release of the long-range morphogen Dpp. One could imagine a sequence of such indirect effects so that a cell has to make only a binary response to its neighbour before handing on another signal. Genetic or embryological mosaics of cells carrying dominant

negative or mutationally defective receptors have shown, at least for the *Drosophila* wing²⁵, *Xenopus* embryo¹⁶ and fish embryo¹⁵, that near and distant cells respond directly to the same morphogen.

Perception at the cell surface

A cell may respond to morphogen concentration through its receptors in two ways. One is to be armed with receptors having different binding characteristics; for example, high- and low-affinity receptors and their transduction pathways could operate at low and high concentrations of ligand, respectively. The other is to vary the occupancy of one type of receptor, and hence its signalling activity, according to ligand concentration. We therefore need to know whether different morphogen responses are transmitted by one or more kinds of receptor. The *Drosophila* morphogen Dpp is bound initially by the type II receptor Punt, which phosphorylates type I receptors Saxophone (Sax) and Tkv. Overexpression of Tkv can substitute for Sax, and increasing amounts of activated Tkv induce the full range of concentration-dependent cell fates^{26,27}. It seems that the two type I receptors use the same intracellular transduction pathway, Sax being used in normal development to amplify the effects of Tkv. The specificity of these receptors has also been examined by the effects of dominant negative variants; dominant negative Tkv blocks all activity of Dpp, and it is clear that Dpp alone can specify all positional values in a concentration-dependent way⁶. The receptors for *Drosophila*

Wg are of two kinds, with a tenfold difference in ligand affinity measured *in vitro*²⁸, but this difference is not related to Wg concentration response in the wing, and these receptors are functionally redundant²⁹.

Vertebrate TGF- β factors including activin are also bound initially by a type II receptor; this phosphorylates an associated type I receptor, which in turn phosphorylates Smad2, causing it to enter the nucleus. From binding studies on isolated cells from the *Xenopus* animal cap³⁰, activin binding is known to be limited only by type II receptors, at all concentrations, and that these receptors have only one class of affinity, judged by Scatchard analysis. The inactivation of type I activin receptors by dominant negative overexpression eliminates response to all concentrations of activin, showing that these same receptors are necessary and sufficient components of the pathway, at all concentrations³¹. All of these results indicate that cells read different concentrations of a morphogen by varying the activity of one type of receptor and not by making use of different receptors for the same morphogen.

We need eventually to understand morphogen perception in terms of receptor occupancy. To understand how a cell can measure morphogen concentration using only one kind of receptor, we first need to know whether it is measuring the absolute number of occupied receptors or a proportion of these, such as the ratio of occupied to unoccupied receptors. This last idea could operate if, for example, unoccupied receptors have a phosphatase activity that counteracts the known serine-threonine kinase activity of TGF- β type II receptors. The overexpression, by up to tenfold, of the activin type II receptor in *Xenopus* embryo cells shows that the choice of gene response depends on the absolute number of occupied receptors, entirely independently of how many unoccupied receptors are present³⁰.

An understanding of receptor signalling with time would be complicated if receptors turn over or are recycled rapidly. In the case of activin, it seems that receptors bind their ligand with high affinity so that it remains bound for several hours³⁰. Over this timescale, receptors and their bound ligand are stable and seem not to be internalized. This situation is certainly not universal. Many receptors are internalized and downregulated soon after ligand binding. In the particular case of major histocompatibility complex (MHC) receptors, these continue to signal after the ligand, which binds with low affinity, has dissociated^{32,33}. For morphogen interpretation that takes many hours or days (such as in the vertebrate limb), receptor turnover must certainly be taken into account.

From receptor to gene

We need next to know at what stage in the pathway from receptor to gene there is a change from a quantitative difference in the number of occupied receptors to a qualitative choice of gene expression. In the rather few cases so far analysed, this switch does not happen in the known transduction pathways. Thus Armadillo is necessary and sufficient for the induced expression of genes that respond to both short- and long-range Wg signalling³⁴. Likewise, Smad2 is required for all responses to activin signalling. Furthermore, the threefold difference in extracellular activin concentration that causes a switch in *Xbra/gsc* gene expression is accompanied by only a threefold increase in the level of nuclear Smad2 concentration^{35,36}. Thus, surprisingly, morphogen concentration seems to be transmitted from outside the cell to the nucleus by a single pathway with no amplification.

The examination of activated transduction molecules is proving valuable in visualizing the effects of morphogen gradients. Antibodies against activated (phosphorylated) Smad1 reveal a gradient distribution with a high point on the ventral side of the embryo, as expected if this reflects the BMP gradient³⁷. Over- and under-expression of BMP and activin changes the gradient distribution, as predicted for Smad1 and Smad2 (ref. 35). Likewise, a gradient of phosphorylated Mad is seen in *Drosophila* embryos with a high

point in the most dorsal region of the blastoderm embryo, and this is influenced in accord with an up- or downregulation of Dpp and Screw morphogen expression³⁸. The direct demonstration of these gradients of activated TGF- β transduction molecules provides valuable support for the existence of morphogen gradients in embryos, as these presumed gradients are very hard to see directly.

The next step for the determination of the threshold responses takes place at the level of transcription. In no case so far has the mechanism by which changes in extracellular morphogen concentration activate different genes at the transcriptional level been exactly deduced. Several ideas exist, however, and two of these are presented in Fig. 3.

Kinetics of interpretation

A proper understanding of morphogen action requires a knowledge of timing. An interesting 'sequential cell context' hypothesis proposes that a low dose of morphogen first activates a 'low-concentration' response, which, at a later time, induces a 'high-concentration' response³⁹. The difficulty with this idea seems to be that the subsequent response is indirect, and this model cannot therefore explain those morphogen responses known to be direct

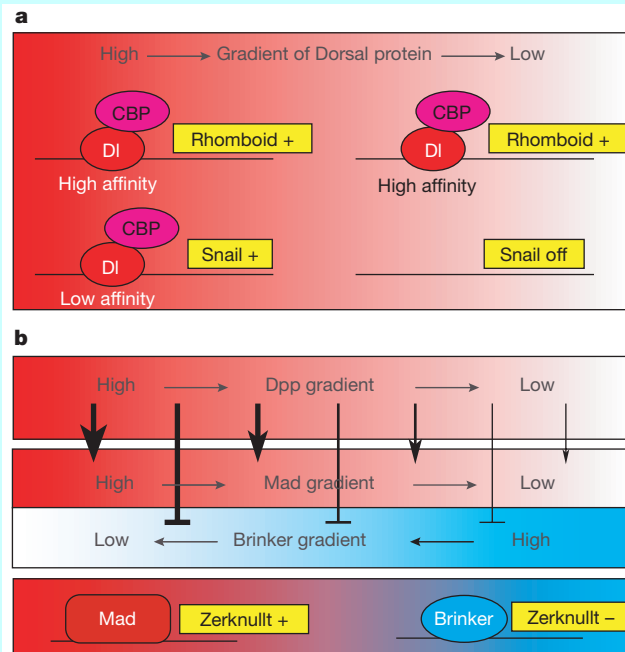


Figure 3 Two possible mechanisms of concentration-dependent transcription. Dorsal (DI) is a maternal cytoplasmic protein that moves to the nucleus of ventral cells in the *Drosophila* embryo, and activates the promoters of ventral (mesodermal) genes. It is present in nuclei in a gradient from high (ventral) to low (dorsal) concentration, opposite to the gradient of Dpp (Fig. 1a). **a**, Promoter binding affinity. In the *Drosophila* embryo, Snail (mesoderm) has weak binding sites that respond only to high concentrations of DI, whereas Rhomboid has high-affinity binding sites for DI and so is expressed at both high and low concentrations of DI⁴⁷. **b**, Competition between activator and repressor. In *Drosophila* Dpp signalling, Brinker, the transcription of which is repressed by Dpp, acts as a repressor of Dpp target-gene transcription. Several Mad- and Brinker-binding sites have been described in the promoter of *zerknullt* (*zen*) and many of these overlap. In the early embryo, competition between these two factors determines the expression domain of *zen*. In cells responding to a high concentration of Dpp, the high concentration of phosphorylated Mad outcompetes Brinker for binding to the promoter of *zen*, which is therefore expressed. In the cells receiving a lower concentration of Dpp, Brinker, which is not downregulated, represses *zen* expression by preventing the binding of Mad and by recruiting co-repressor to the promoter⁴⁹.

(cycloheximide insensitive). Here we discuss three ideas on how cells make direct responses to morphogen gradients.

Our first proposition is that the availability of ligand (morphogen) is the limiting factor in determining the level of response to concentration. In theory any component of a pathway, such as the availability of receptors, the supply of transduction molecules, and so on, could prevent or restrict the ability of a cell to recognize the concentration of an extracellular molecule. For example, if all of a cell's receptors were occupied with a low morphogen concentration, the cell could not perceive a higher concentration until more receptors had been synthesized, and the availability of these would determine the ability of a cell to know its position in a gradient. The following results indicate that ligand supply is usually limiting. In the *Drosophila* wing disc, twofold increases in Dpp extend the range of concentration-related responses⁶; similar increases in receptor abundance have little effect. Dpp represses the level of its receptor Tkv⁴⁰, so that Dpp is probably not limiting near its source, and both response genes are activated equally during the first 50 h. However, further from the Dpp source, receptor levels are higher, and it is presumed that Dpp is limiting in this more distant region where cells read the gradient to differentially activate *sal* and *omb*. Also, in the wing disc, Wg downregulates its receptor Frizzled (Fz), so that it is in lower abundance near the source than further from it⁴¹. Ptc absorbs Hh and Sonic hedgehog (Shh) near their source, thereby making their supply limiting further away^{24,42,43}. We know in *Xenopus* that small increases in activin concentration result in stepped changes in the types of gene expressed⁴⁴, as expected if morphogen concentration is limiting. We also know that an increase in receptor abundance does not change gene response³⁰. Moderate increases in Smad2 supply do not affect response, although massive increases activate the signalling pathway in the absence of ligand. Because less than 10% of the activin receptors are occupied when gene switches at low concentration take place, it is hard to understand why the available supply of ligand is not all absorbed by the cells with empty receptors near the signalling source. This apparent inconsistency can be explained, however, by supposing that most of the morphogen molecules remain in the intercellular space between cells, either free or attached with low affinity and high capacity to the extracellular matrix³⁰. Although activin can be bound rapidly at a high concentration (1–4 nM) in dissociated cells, at *in vivo* concentrations (50–200 pM), the time required for binding is long and should permit unbound ligand molecules to pass out of reach of receptors on cells near the source.

Our second proposition is that cells respond to ligand concentration according to the absolute number of receptors occupied at any time. Most importantly, there is no integration of the number of occupied receptors multiplied by time. Thus a given number of receptors occupied for 1, 2 or 4 h will all elicit the same choice of response, even though a two or four times greater number of occupied receptors induces qualitatively different gene responses. After only a 10-min exposure to activin, the number of occupied receptors in dissociated cells from the *Xenopus* animal cap remains constant for a few hours, at, for example, 300 of a total 5,000 receptors per cell. Under these conditions, gene expression does not change, whether tested 2 or 4 h later³⁰. This is entirely consistent with the complementary observation that there is indeed an integration of receptor occupancy and time as cells fill receptors with available surrounding morphogen: a cell in 4 nM activin will fill twice as many receptors in 20 min as in 10 min, and twice as many in 8 nM activin as in 4 nM activin for 10 min. This principle has also been demonstrated by varying the duration of signalling by the Nodal cofactor Oep in zebrafish axis formation⁴⁵. We therefore propose that the loading of receptors by ligand increases progressively with time, but that, once loaded, the activity of a receptor remains constant with time.

Our third proposition concerns gene read-out. We suggest that a

cell with a particular number of occupied receptors will continue to express the same gene until either the occupancy of receptors goes up or the period of competence terminates. Bead exchange experiments have indicated a ratchet effect by which a cell can rapidly change its response to morphogen concentration in an upward direction, by filling a higher proportion of its receptors, but can change downwards only very slowly, as receptors become vacated. In the particular case of the gene *Xbra*, it is transcribed less strongly as activin concentration goes up and higher response genes such as *gsc* or *eomes* are expressed. This is, however, an indirect effect¹⁸, and much more generally a low-concentration-response gene such as *omb* in the *Drosophila* wing disc continues to be expressed when a high response gene such as *sal* is activated. Likewise, *neuralized* (*neur*), *Distalless* (*Dll*) and *vestigial* (*vg*) are all transcribed at the highest Wg concentrations. It is not yet clear how the co-expression of two or more morphogen response genes at the same time, and presumably in the same cells, generates the appropriate morphological or developmental effect.

Finally, we comment on a particularly intriguing future problem in the interpretation of morphogen gradients: how cells respond to an extracellular concentration that changes with time. In development, signalling centres start to emit their signalling factor or morphogen at a particular stage. On the assumption that the supply of morphogen is limiting (above), cells nearest the source will at first experience a low concentration of the morphogen, but this will increase until it reaches the much higher level to which cells near the source make their correct response. We want to know how these cells near the source know when to interpret the factor concentration around them. There are two ways in which cells

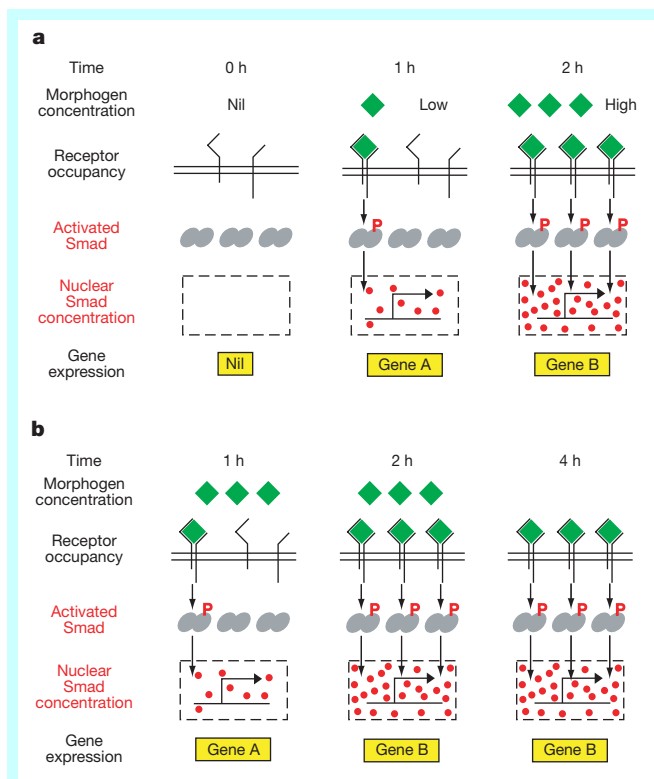


Figure 4 Concepts of morphogen gradient interpretation. **a**, Morphogen concentration that increases with time. The model proposes that the number of occupied receptors, and of phosphorylated Smad molecules, increases with time. **b**, Morphogen concentration that is constant or declines with time. The model proposes that, with constant morphogen concentration, receptor occupancy and signalling activity increases. With a constant number of occupied receptors, it is proposed that the level of intracellular signalling activity also remains constant.

might solve this problem. One is to wait until a steady-state gradient has been reached and respond only at that time; the other is to respond continuously to their ambient concentration and to change their gene response accordingly with time. The latter is more likely, because it is known from bead exchange experiments that cells can change their expression upwards but not downwards during the course of an activin response⁴⁶. Our current ideas on how cells interpret gradient concentrations that change with time are summarized in Fig. 4. To provide clear answers to these and other questions, future work should endeavour to monitor the transduction, and perhaps even gene transcription, response to morphogen interpretation in single cells in real time. □

1. Wolpert, L. Positional information revisited. *Development (Suppl.)* 3–12 (1989).
2. Cooke, J. Morphogens in vertebrate development: how do they work? *BioEssays* 17, 93–96 (1995).
3. Lawrence, P. A. & Struhl, G. Morphogens, compartments, and pattern: lessons from *Drosophila*? *Cell* 85, 951–961 (1996).
4. Neumann, C. & Cohen, S. Morphogens and pattern formation. *BioEssays* 19, 721–729 (1997).
5. Teleman, A. A., Strigini, M. & Cohen, S. M. Shaping morphogen gradients. *Cell* 105, 559–562 (2001).
6. Podos, S. D. & Ferguson, E. L. Morphogen gradients: new insights from DPP. *Trends Genet.* 15, 396–402 (1999).
7. De Robertis, E. M., Larraín, J., Oelgeschläger, M. & Wessely, O. The establishment of Spemann's organizer and patterning of the vertebrate embryo. *Nature Rev. Genet.* 1, 171–181 (2000).
8. Driever, W. & Nüsslein-Volhard, C. The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* 54, 95–104 (1988).
9. Entchev, E. V., Schwabedissen, A. & Gonzalez-Gaitan, M. Gradient formation of the TGF- β homolog Dpp. *Cell* 103, 981–991 (2000).
10. Teleman, A. A. & Cohen, S. M. Dpp gradient formation in the *Drosophila* wing imaginal disc. *Cell* 103, 971–980 (2000).
11. Dubois, L., Lecourtis, M., Alexandre, C., Hirst, E. & Vincent, J. P. Regulated endocytic routing modulates wingless signaling in *Drosophila* embryos. *Cell* 105, 613–624 (2001).
12. Hogan, B. L., Thaller, C. & Eichele, G. Evidence that Hensen's node is a site of retinoic acid synthesis. *ature* 359, 237–241 (1992).
13. Slack, J. M. W. Inducing factors in *Xenopus* early embryos. *Curr. Biol.* 4, 116–126 (1994).
14. Porter, J. A., Young, K. E. & Beachy, P. A. Cholesterol modification of hedgehog signaling proteins in animal development. *Science* 274, 255–259 (1996).
15. Chen, Y. & Schier, A. F. The zebrafish Nodal signal Squint functions as a morphogen. *Nature* 411, 607–610 (2001).
16. McDowell, N., Zorn, A. M., Crease, D. J. & Gurdon, J. B. Activin has a direct long range signalling activity and can form a concentration gradient by diffusion. *Curr. Biol.* 7, 671–681 (1997).
17. Gurdon, J. B., Harger, P., Mitchell, A. & Lemaire, P. Activin signalling and response to a morphogen gradient. *Nature* 371, 487–492 (1994).
18. Papin, C. & Smith, J. C. Gradual refinement of activin-induced thresholds requires protein synthesis. *Dev. Biol.* 217, 166–172 (2000).
19. Ferguson, E. L. & Anderson, K. V. Decapentaplegic acts as a morphogen to organize dorsal–ventral pattern in the *Drosophila* embryo. *Cell* 71, 451–461 (1992).
20. Lecuit, T. *et al.* Two distinct mechanisms for long-range patterning by Decapentaplegic in the *Drosophila* wing. *Nature* 381, 387–393 (1996).
21. Wharton, K. A., Ray, R. P. & Gelbart, W. M. An activity gradient of decapentaplegic is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo. *Development* 117, 807–822 (1993).
22. Wolpert, L. *Principles of Development* 399–401 (Oxford Univ. Press, Oxford, 1998).
23. Gurdon, J. B. *et al.* Single cells can sense their position in a morphogen gradient. *Development* 126, 5309–5317 (1999).
24. Chen, Y. & Struhl, G. Dual roles for patched in sequestering and transducing Hedgehog. *Cell* 87, 553–563 (1996).
25. Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range action of a DPP morphogen gradient. *Cell* 85, 357–368 (1996).
26. Neul, J. L. & Ferguson, E. L. Spatially restricted activation of the SAX receptor by SCW modulates DPP/TKV signaling in *Drosophila* dorsal–ventral patterning. *Cell* 95, 483–494 (1998).
27. Nguyen, M., Park, S., Marques, G. & Arora, K. Interpretation of a BMP activity gradient in *Drosophila* embryos depends on synergistic signaling by two type I receptors, SAX and TKV. *Cell* 95, 495–506 (1998).
28. Rulifson, E. J., Wu, C. H. & Nusse, R. Pathway specificity by the bifunctional receptor frizzled is determined by affinity for wingless. *Mol. Cell* 6, 117–126 (2000).
29. Chen, C. M. & Struhl, G. Wingless transduction by the Frizzled and Frizzled2 proteins of *Drosophila*. *Development* 126, 5441–5452 (1999).
30. Dyson, S. & Gurdon, J. B. The interpretation of position in a morphogen gradient as revealed by occupancy of activin receptors. *Cell* 93, 557–568 (1998).
31. Hemmati-Brivanlou, A. & Melton, D. A. A truncated activin receptor inhibits mesoderm induction and formation of axial structures in *Xenopus* embryos. *Nature* 359, 609–614 (1992).
32. Valitutti, S., Muller, S., Cella, M., Padovan, E. & Lanzavecchia, A. Serial triggering of many T-cell receptors by a few peptide–MHC complexes. *Nature* 375, 148–151 (1995).
33. Gurdon, J. B., Dyson, S. & St Johnston, D. Cells' perception of position in a concentration gradient. *Cell* 95, 159–162 (1998).
34. Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of a wingless morphogen gradient. *Cell* 87, 833–844 (1996).
35. Graff, J. M., Bansal, A. & Melton, D. A. *Xenopus* mad proteins transduce distinct subsets of signals for the TGF β superfamily. *Cell* 85, 479–487 (1996).
36. Shimizu, K. & Gurdon, J. B. A quantitative analysis of signal transduction from activin receptor to nucleus and its relevance to morphogen gradient interpretation. *Proc. Natl Acad. Sci. USA* 96, 6791–6796 (1999).
37. Faure, S., Lee, M. A., Keller, T., ten Dijke, P. & Whitman, M. Endogenous patterns of TGF β superfamily signaling during early *Xenopus* development. *Development* 127, 2917–2931 (2000).
38. Dorfman, R. & Shilo, B. Z. Biphasic activation of the BMP pathway patterns the *Drosophila* embryonic dorsal region. *Development* 128, 965–972 (2001).
39. Pages, F. & Kerridge, S. Morphogen gradients. A question of time or concentration? *Trends Genet.* 16, 40–44 (2000).
40. Lecuit, T. & Cohen, S. M. Dpp receptor levels contribute to shaping the Dpp morphogen gradient in the *Drosophila* wing imaginal disc. *Development* 125, 4901–4907 (1998).
41. Cadigan, K. M., Fish, M. P., Rulifson, E. J. & Nusse, R. Wingless repression of *Drosophila* frizzled 2 expression shapes the Wingless morphogen gradient in the wing. *Cell* 93, 767–777 (1998).
42. Jessell, T. M. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature Rev. Genet.* 1, 20–29 (2000).
43. Briscoe, J., Chen, Y., Jessell, T. M. & Struhl, G. A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol. Cell* 7, 1279–1291 (2001).
44. Green, J. B., New, H. V. & Smith, J. C. Responses of embryonic *Xenopus* cells to activin and FGF are separated by multiple dose thresholds and correspond to distinct axes of the mesoderm. *Cell* 71, 731–739 (1992).
45. Gritsman, K., Talbot, W. S. & Schier, A. F. Nodal signaling patterns the organizer. *Development* 127, 921–932 (2000).
46. Gurdon, J. B., Mitchell, A. & Mahony, D. Direct and continuous assessment by cells of their position in a morphogen gradient. *Nature* 376, 520–521 (1995).
47. Mannervik, M., Nibu, Y., Zhang, H. & Levine, M. Transcriptional coregulators in development. *Science* 284, 606–609 (1999).
48. Ashe, H. L., Mannervik, M. & Levine, M. Dpp signaling thresholds in the dorsal ectoderm of the *Drosophila* embryo. *Development* 127, 3305–3312 (2000).
49. Affolter, M., Marty, T., Vigano, M. A. & Jazwinska, A. Nuclear interpretation of Dpp signaling in *Drosophila*. *EMBO J.* 20, 3298–3305 (2001).
50. Klein, T. & Arias, A. M. The vestigial gene product provides a molecular context for the interpretation of signals during the development of the wing in *Drosophila*. *Development* 126, 913–925 (1999).

Acknowledgements

We apologize to the authors of original work not cited. Owing to space limitations, we cited review articles where possible. Our work is supported by the Cancer Research Campaign and the European Community.

Correspondence should be addressed to J.B.G. (e-mail: j.gurdon@welc.cam.ac.uk).

Locally critical quantum phase transitions in strongly correlated metals

Qimiao Si*, Silvio Rabello*, Kevin Ingersent† & J. Llewellyn Smith*

* Department of Physics & Astronomy, Rice University, Houston, Texas 77251-1892, USA

† Department of Physics, University of Florida, Gainesville, Florida 32611-8440, USA

When a metal undergoes a continuous quantum phase transition, non-Fermi-liquid behaviour arises near the critical point. All the low-energy degrees of freedom induced by quantum criticality are usually assumed to be spatially extended, corresponding to long-wavelength fluctuations of the order parameter. But this picture has been contradicted by the results of recent experiments on a prototype system: heavy fermion metals at a zero-temperature magnetic transition. In particular, neutron scattering from $\text{CeCu}_{6-x}\text{Au}_x$ has revealed anomalous dynamics at atomic length scales, leading to much debate as to the fate of the local moments in the quantum-critical regime. Here we report our theoretical finding of a locally critical quantum phase transition in a model of heavy fermions. The dynamics at the critical point are in agreement with experiment. We propose local criticality to be a phenomenon of general relevance to strongly correlated metals.

Quantum (zero-temperature) phase transitions are ubiquitous in strongly correlated metals (for a recent review, see ref. 1). The extensive present interest in metals close to a second-order quantum phase transition has stemmed largely from studies of high-temperature superconductors. In these systems, however, it has been hard to locate the putative quantum critical points (QCPs). The situation seems to be simpler in some related families of strongly correlated metals. In particular, there are many heavy-fermion metals that can be tuned between an antiferromagnetic metal and a paramagnetic metal. In recent years, QCPs have been explicitly identified in a number of stoichiometric or nearly stoichiometric systems. Non-Fermi-liquid behaviour² is observed, usually through transport and thermodynamic measurements, in the quantum-critical regime. Examples^{3–11} include CePd_2Si_2 , $\text{CeCu}_{6-x}\text{Au}_x$, CeNi_2Ge_2 and YbRh_2Si_2 .

The most fundamental question regarding any condensed matter system concerns the nature of its low-energy excitations. The traditional theory for metals close to a QCP assumes that the only important degrees of freedom are long-wavelength fluctuations of the order parameter. For a magnetic QCP, such long-wavelength fluctuations are usually called paramagnons. This theory has its origins back in the 1960s in the description of liquid ^3He (for a historical perspective, see ref. 12), and was subsequently formulated in the modern language of critical phenomena^{13,14}. The assumption that the long-wavelength paramagnons are the only critical degrees of freedom, adopted from the theory of classical (finite-temperature) phase transitions, turns out to have three important consequences. First, the dynamical spin susceptibility at generic wavevectors (wavevectors far from the antiferromagnetic ordering wavevector) has the usual Fermi-liquid form, that is, it is linear in frequency. Second, even close to the ordering wavevector, the dynamical spin susceptibility is also linear in frequency. Third, there is a violation of the so-called ω/T scaling (defined in the next section).

This picture has been subjected to experimental testing in recent years. Heavy-fermion metals have undergone by far the most systematic study. A major puzzle has been raised by neutron-scattering experiments^{9–11} on $\text{CeCu}_{6-x}\text{Au}_x$, which goes from a paramagnetic metal to an antiferromagnetic metal as x increases through a critical value, $x_c \approx 0.1$ (ref. 4). Close to the antiferromagnetic ordering wavevector, a fractional exponent ($\alpha < 1$) appears in the frequency and temperature dependences of the dynamical spin susceptibility. The same exponent α describes the frequency and temperature dependences at generic wavevectors,

over essentially the entire Brillouin zone. Finally, the dynamical spin susceptibility exhibits ω/T scaling. These experiments have led to much debate as to the nature of the QCPs in heavy-fermion metals, especially concerning the fate of the local magnetic moments in the quantum-critical regime^{9–11,15–18}.

Here we show that a dynamical spin susceptibility of the form observed experimentally may arise at a locally critical quantum phase transition, where local critical degrees of freedom coexist with spatially extended ones. Although we reach this conclusion by solving a microscopic model appropriate for heavy-fermion metals, the possibility of such a QCP can be seen on general grounds. The universality class of a classical phase transition is entirely determined by statics¹. The critical degrees of freedom must be spatially extended, as only long-wavelength fluctuations of the order parameter are energetically favourable. At a quantum phase transition, by contrast, both the statics and dynamics (that is, quantum fluctuations) are important¹. The effect of quantum fluctuations on local degrees of freedom is to introduce a coupling to certain dissipative baths. When this coupling is sufficiently strong, the local degrees of freedom can become critical. (A similar situation—local degrees of freedom driven critical by a dissipative bath—arises in macroscopic quantum tunnelling problems¹⁹.)

Locally critical quantum phase transition

The microscopic model that we consider here is the Kondo lattice model²⁰ (Fig. 1a). At each lattice site, a local moment²¹ interacts via an exchange coupling J_K with the spin of any conduction electron sitting at the site. There are two important energy scales in the problem^{22,23}: the Kondo temperature T_K sets the scale below which an isolated local moment would be screened by the spins of the conduction electrons, while the Ruderman–Kittel–Kasuya–Yosida (RKKY) interaction characterizes the induced coupling between two local moments. We will assume an average occupancy of less than one conduction electron per lattice site; as a result, all phases considered here are metallic. To systematically analyse the behaviour of the local degrees of freedom in this model, we apply the extended dynamical mean-field theory (EDMFT) developed in refs 24–26. As in the standard dynamical mean-field theory²⁷, the correlation functions of the lattice problem in the EDMFT are calculated through a self-consistent impurity Kondo problem. The latter (Fig. 1b) describes a local moment simultaneously coupled to two dissipative baths¹⁹. A fermionic bath accounts for all temporal fluctuations arising from hopping of electrons between the local site

and the rest of the lattice, while a bosonic bath represents the fluctuating magnetic field generated by the local moments at all other sites. The couplings to the two baths are J_K and g , respectively. Further details are given in the next section.

We find two types of QCP, illustrated by the phase diagrams in Figs 2 and 3. In each case, the tuning parameter δ is the ratio of the RKKY interaction to the Kondo temperature. Increasing δ has two effects. First, the magnetic correlations become more pronounced. The dynamical spin susceptibility $\chi(\mathbf{Q}, \omega)$ —where $\mathbf{Q}$ is the peak wavevector—diverges at some threshold value $\delta = \delta_c$, which defines a QCP separating a paramagnetic metal ($\delta < \delta_c$) from an antiferromagnetic metal ($\delta > \delta_c$). Second, the local Kondo physics governed by the effective impurity problem is also changed^{28–30}. Through self-consistency, an increase in δ causes an increase in the coupling g between the local moment and the fluctuating magnetic field. At some value $\delta = \delta_{loc}^c$, the corresponding ratio g/T_K reaches a threshold value where the local Kondo problem becomes critical. The existence of such a critical point reflects two competing processes in the local problem (Fig. 1b): quenching of the local moment through its Kondo coupling to the spins of the fermionic bath, and the tendency of the coupling g to polarize the local moment along the direction of the fluctuating magnetic field. At $\delta = \delta_{loc}^c$, the fluctuating magnetic field just succeeds in preventing the spins of the fermionic bath from completely quenching the local moment, yielding a singular local susceptibility. This critical point is also marked by the vanishing of a local energy scale E_{loc}^* , reflecting a critical slowing down. E_{loc}^* serves as an effective Fermi energy scale, in the sense that the local susceptibility has the usual Fermi-liquid form (that is, an imaginary part that is linear in the frequency ω) only at energies below E_{loc}^* .

At the first type of QCP (Fig. 2), the lattice system orders before the effective local problem has a chance to become critical, that is, $\delta_c < \delta_{loc}^c$, so E_{loc}^* remains finite at the transition. The local moments are completely quenched at zero temperature over the entire paramagnetic region and also at the QCP. Their only vestiges are

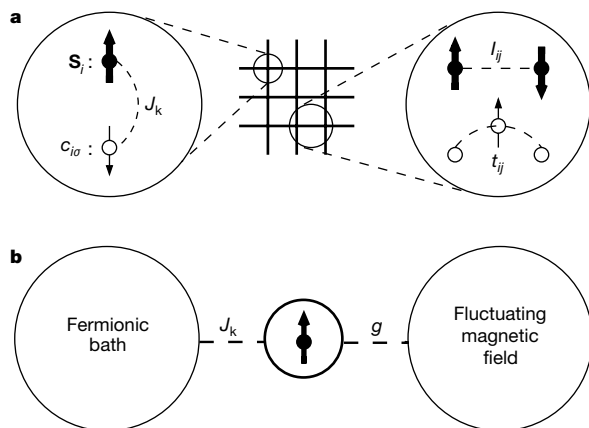


Figure 1 A theoretical model of heavy fermions. **a**, The Kondo lattice model. At each lattice site i , a spin-1/2 local moment $\mathbf{S}_i$ interacts with a local conduction electron orbital c_{ia} through an antiferromagnetic Kondo coupling of strength J_K . The hopping amplitude between the conduction electron orbitals at sites i and j is t_{ij} . The Fourier transform of t_{ij} gives the band dispersion $\epsilon_{\mathbf{k}}$, corresponding to which is a density of states $\rho_0(\epsilon)$. The Kondo temperature T_K is the characteristic scale for the screening of an isolated local moment by the spins of the conduction electrons. For sufficiently small J_K , $T_K \approx [\rho_0(\mu)]^{-1} \exp[-1/\rho_0(\mu)J_K]$, where μ is the chemical potential. The RKKY interaction between two local moments at sites i and j is I_{ij} , the Fourier transform of which is $I_{\mathbf{q}}$. **b**, The effective impurity Kondo model to which the Kondo lattice model is mapped in the EDMFT^{24–26}. The local moment is coupled to two dissipative baths, one fermionic and the other bosonic. The bosonic bath describes a fluctuating magnetic field. The couplings to the two baths are J_K and g , respectively. The energy dispersions of the two baths are determined self-consistently.

the Kondo resonances, which hybridize with the conduction electrons to form Landau quasiparticles^{20,22,23,31}. The RKKY interaction manifests itself as an interaction between these quasiparticles, leading to a spin-density-wave instability which drives the transition²². As a result, the dynamical spin susceptibility has the form as in the traditional theory^{13,14}. We find this type of QCP when the spin fluctuations are three-dimensional.

The second type of QCP is illustrated in Fig. 3. Here, $\delta_c = \delta_{loc}^c$, so the lattice system reaches its ordering transition precisely at the point where the local problem also becomes critical. Thus, E_{loc}^* vanishes at the QCP, and two kinds of critical degrees of freedom coexist: long-wavelength fluctuations of the order parameter, and local fluctuations originating from the local moments. The transition is locally critical. We find this type of QCP when spin fluctuations are two-dimensional. (In order for the ordering temperature T_N shown in Fig. 3 to be nonzero for $\delta > \delta_c$, it is necessary to have an infinitesimal RKKY coupling in the third dimension.)

At the locally critical point, the zero-temperature dynamical spin susceptibility has an anomalous frequency dependence at wavevectors $\mathbf{q}$ not only close to the ordering wavevector $\mathbf{Q}$ but everywhere else in the Brillouin zone as well:

$$\chi(\mathbf{q}, \omega) = \frac{1}{f(\mathbf{q}) + A\omega^\alpha} \quad (1)$$

Here α is an anomalous exponent (an expression for which is given in equation (13) below). For any wavevector $\mathbf{q}$ far away from $\mathbf{Q}$, $f(\mathbf{q})$ is non-zero and varies smoothly with $\mathbf{q}$; for $\mathbf{q}$ close to $\mathbf{Q}$, $f(\mathbf{q})$ is proportional to $(\mathbf{q} - \mathbf{Q})^2$.

At finite temperatures, ω^α in equation (1) is replaced by $T^\alpha W(\omega/T)$, where $W(\omega/T)$ is a scaling function whose form depends only on α . This has two experimentally testable consequences. First, the dynamical spin susceptibility at $\mathbf{q} = \mathbf{Q}$ satisfies an ω/T scaling:

$$\chi(\mathbf{Q}, \omega, T) = \frac{1}{AT^\alpha W(\omega/T)} \quad (2)$$

Second, the static uniform spin susceptibility has a modified Curie-Weiss form

$$\chi(T) = \frac{1}{\Theta + BT^\alpha} \quad (3)$$

where Θ is a positive constant.

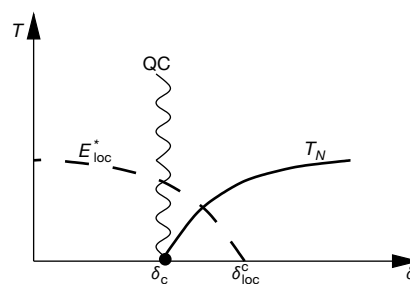


Figure 2 Diagram of a conventional quantum phase transition in Kondo lattices. The tuning parameter δ is the ratio of the RKKY interaction to the Kondo energy scale T_K . When δ is small the local moments are completely screened by the spins of the conduction electrons, and the system is a paramagnetic metal. Increasing δ strengthens the coupling between each local moment and the fluctuating magnetic field produced by all other local moments, as measured by the ratio g/T_K , where g parametrizes the effective local problem defined in Fig. 1b and equation (6). At δ_{loc}^c , the corresponding g/T_K would reach the critical value for a phase transition in the local dissipative problem (see main text); at this point, a local energy scale E_{loc}^* (dashed line) would vanish. Increasing δ also increases the spin susceptibility at the peak wavevector, $\chi(\mathbf{Q})$, which diverges at a quantum critical point at $\delta = \delta_c$. Beyond δ_c , there exists a finite transition temperature T_N (solid line) below which the system is an antiferromagnetic metal. In this conventional case, $\delta_c < \delta_{loc}^c$, that is, E_{loc}^* is finite at the critical point. The quantum-critical (QC) regime, denoted by the wavy line, is described by the traditional theory.

Our results can be directly compared to neutron-scattering experiments on Au-doped $\text{CeCu}_{6-x}\text{Au}_x$ at the critical concentration $x_c \approx 0.1$. (1) The experimental data have been fitted^{9,11} to the form of equation (1), with an exponent $\alpha \approx 0.75$. (2) Within experimental resolution, $f(\mathbf{q})$ goes to zero along lines in the three-dimensional Brillouin zone, implying that the magnetic fluctuations are two-dimensional in real space^{10,11,32}. (3) An ω/T scaling has been extensively reported^{9,11} (see also ref. 33).

In addition, the modified Curie–Weiss form (equation (3)) is known^{9,11} to fit the magnetization data in $\text{CeCu}_{6-x}\text{Au}_x$. This form also appears to describe the magnetization of some other heavy-fermion metals near quantum criticality⁸. One example (F. Steglich, personal communication) is YbRh_2Si_2 , which is already very close to being critical at ambient pressure and without any doping. The thermodynamic and transport properties⁵ of this compound are very similar to those of $\text{CeCu}_{6-x}\text{Au}_x$ at $x = x_c$.

From the local susceptibility we can also determine the NMR spin–lattice relaxation rate $1/T_1$. Barring subtle form-factor cancellations, the relaxation rate is expected to be temperature-independent in the relevant temperature range:

$$\frac{1}{T_1} \approx \text{constant} \quad (4)$$

This prediction could be tested, for instance, through NMR measurements on the Cu sites in $\text{CeCu}_{6-x}\text{Au}_x$ or the Si sites in YbRh_2Si_2 .

Extended dynamical mean-field theory

We now outline the EDMFT analysis leading to the above conclusions. The Kondo lattice model illustrated in Fig. 1a is specified by the hamiltonian

$$H = \sum_{ij,\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \sum_i J_K \mathbf{S}_i \cdot \mathbf{s}_{c,i} + \sum_{ij} I_{ij} \mathbf{S}_i \cdot \mathbf{S}_j \quad (5)$$

where $\mathbf{s}_{c,i}$ is the conduction electron spin at site i . Within the EDMFT^{24–26}, equation (5) is mapped onto an effective single-site problem, illustrated in Fig. 1b and represented by the hamiltonian

$$H_{\text{loc}} = J_K \mathbf{S} \cdot \mathbf{s}_c + \sum_{p,\sigma} E_p c_{p\sigma}^\dagger c_{p\sigma} + g \sum_p \mathbf{S} \cdot (\phi_p + \phi_p^\dagger) + \sum_p w_p \phi_p^\dagger \phi_p \quad (6)$$

where $c_{p\sigma}$ and ϕ_p describe the fermionic and bosonic dissipative baths, respectively. The self-consistent procedure is as follows. We put in trial forms for E_p , w_p and g , and solve the impurity Kondo problem to determine the electron self-energy $\Sigma(\omega)$, the ‘spin self-energy’ $M(\omega)$, the local Green’s function $G_{\text{loc}}(\omega)$, and the local susceptibility $\chi_{\text{loc}}(\omega)$. (For definitions of these functions and further details of the calculational procedure, see Methods. We note here that the key approximation of the EDMFT is the assumption that the electron and spin self-energies are momentum-independent.)

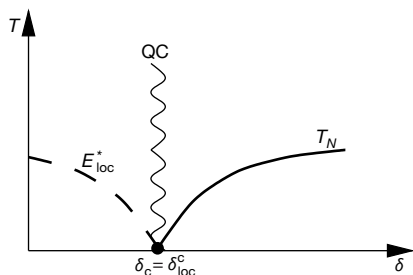


Figure 3 Diagram of a locally critical quantum phase transition in Kondo lattices. The notation is the same as in Fig. 2. The crucial new feature here is that $\delta_c = \delta_c^{\text{loc}}$, that is, E_{loc}^* vanishes at the quantum critical point. Local and spatially-extended critical degrees of freedom coexist in the quantum-critical (QC) regime.

Self-consistency is imposed by demanding that G_{loc} and χ_{loc} are equal to the wavevector averages, respectively, of the lattice Green’s function $G(\mathbf{k}, \omega)$ and the lattice susceptibility

$$\chi(\mathbf{q}, \omega) = \frac{1}{M(\omega) + I_q} \quad (7)$$

where I_q is the Fourier transform of I_{ij} . For the purpose of determining the universal low-energy behaviour, we take the density of states of the fermionic bath near the Fermi energy to be a non-zero constant, and the spectral function of the fluctuating magnetic field to have a power-law dependence on frequency, with an exponent γ , for ω below some cut-off scale Λ . The effective impurity Kondo problem is solved using a $1 - \gamma$ expansion^{28,29,34}. Two types of self-consistent solution are obtained for different forms of the ‘RKKY density of states’:

$$\rho_l(\epsilon) = \sum_q \delta(\epsilon - I_q) \quad (8)$$

The first type of solution (Fig. 2) arises when the RKKY density of states $\rho_l(\epsilon)$ increases from the lower band edge (at $\epsilon = I_Q$) in a square-root fashion (that is, as $\sqrt{\epsilon - I_Q}$). At the QCP, the spin self-energy takes the form

$$M(\omega) = -I_Q - ia\omega \quad (9)$$

where a is a positive, real constant. Through equation (7), this yields a $\chi(\mathbf{q}, \omega)$ of the form given in the traditional theory^{13,14}. The linear frequency dependence of $M(\omega)$ reflects the fact that the spin fluctuations describe quasiparticle–quasihole pairs.

The second type of solution (Fig. 3) arises when $\rho_l(\epsilon)$ increases from zero at the lower band edge with a jump. The self-consistent local dynamical spin susceptibility at the QCP is singular, and is given by

$$\chi_{\text{loc}}(\omega) = \frac{1}{2\Lambda} \ln \frac{\Lambda}{-i\omega} \quad (10)$$

where

$$\Lambda = \frac{2}{\pi \rho_0(\mu)} \exp \left[-\frac{1}{\rho_0(\mu) J_K} \right] \quad (11)$$

$\rho_0(\mu)$ being the bare conduction electron density of states at the chemical potential μ . The corresponding spin self-energy has the form:

$$M(\omega) \approx -I_Q + A\omega^\alpha \quad (12)$$

Here, $A = (-i)^\alpha \Lambda_0 \Lambda^{-\alpha}$, where Λ_0 is defined by the condition that, for $\epsilon \in (I_Q, I_Q + \Lambda_0)$, $\rho_l(\epsilon)$ is approximately equal to its value at the lower edge, $\rho_l(I_Q)$. The exponent α depends on both Λ and $\rho_l(I_Q)$:

$$\alpha = \frac{1}{2\Lambda \rho_l(I_Q)} \quad (13)$$

The spin self-energy acquires this anomalous frequency dependence because spin fluctuations can not only decay into particle–hole pairs, but can also couple to critical local modes. The dynamical spin susceptibility at zero temperature is then given by equation (1), with $f(\mathbf{q}) = I_q - I_Q$. At finite temperatures, the self-consistent solution gives a spin self-energy:

$$M(\omega, T) \approx -I_Q + AT^\alpha W(\omega/T) \quad (14)$$

The scaling function is $W(\omega/T) = (i2\pi)^\alpha \exp[\alpha\psi(1/2 - i\omega/2\pi T)]$, where ψ is the digamma function. This directly leads to equations (2) and (3), with $\Theta = I_{q=0} - I_Q$ and $B \approx \Lambda_0 \Lambda^{-\alpha} (2\pi)^\alpha \exp[\alpha\psi(1/2)]$.

A square-root onset and a jump at the lower edge of the RKKY density of states are characteristic of magnetic fluctuations in three and two dimensions, respectively. These behaviours are exemplified by the RKKY density of states associated with nearest-neighbour coupling in a (three-dimensional) cubic lattice and its counterpart in a (two-dimensional) square lattice. Although the details depend

on the lattice type and the range of interactions, the square-root and jump onsets in the two cases are robust provided that I_q approaches I_0 in a $(\mathbf{q} - \mathbf{Q})^2$ fashion.

Beyond the microscopic theory

So far we have presented the results of a specific microscopic approach: the EDMFT for the Kondo lattice model. We now turn to more general considerations concerning the existence and properties of locally critical quantum phase transitions. There are four main points that we wish to discuss.

First, even in the traditional theory there are hints, albeit very subtle, that two-dimensionality facilitates the realization of a locally critical point. In this theory^{13,14}, the zero-temperature spin fluctuations are given by the harmonic, or gaussian, fluctuations of the order parameter. As a result, the zero-temperature, zero-frequency spin susceptibility is proportional to $(\mathbf{q} - \mathbf{Q})^{-2}$. In two dimensions, the corresponding local—that is, wavevector-averaged—susceptibility is in fact logarithmically singular. (For two-dimensional fermions coupled to two-dimensional commensurate spin fluctuations, there are other kinds of singularities in the spin-density-wave description; see ref. 35.) This singularity can cause important nonlinear effects in the dynamics of any non-trivial local degrees of freedom—such as local moments in heavy-fermion metals—and can ultimately drive them critical. Our EDMFT analysis amounts to a microscopic prescription for these nonlinear effects. We note that the situation is very different for QCPs in insulators^{1,36}. There, the effective field theory is usually below its upper critical dimension, implying that the spin susceptibility is $(\mathbf{q} - \mathbf{Q})^{-2+\eta}$, where the anomalous exponent η takes a positive value. The corresponding local susceptibility is no longer singular in two dimensions.

Second, the contrast between a locally critical transition and a spin-density-wave transition can also be seen at the level of a Ginzburg–Landau-like description. For a spin-density-wave transition, there is only one type of critical degrees of freedom, namely the long-wavelength fluctuations of the order parameter. The Ginzburg–Landau action is then written purely in terms of the frequency-dependent magnetization at wavevectors close to the ordering wavevector:

$$S_{\text{SDW}} = S_{\text{lw}}[\mathbf{m}(\mathbf{q} \approx \mathbf{Q}, \omega)] \quad (15)$$

This generalizes the standard ϕ^4 theory for a classical phase transition in that the fluctuations take place not only in the D spatial dimensions, but also in time. The temporal fluctuations can be thought of as adding z dimensions, where z is the dynamical exponent: S_{lw} is a ϕ^4 theory with an effective dimensionality $D_{\text{eff}} = D + z$ (see refs 1, 13, 14).

For a locally critical point, on the other hand, the Ginzburg–Landau action is no longer just a ϕ^4 theory. There are extra critical modes that characterize the continuous disappearance of the Kondo resonance as the critical point is approached from the paramagnetic side. These modes are independent of the fluctuations of the order parameter. They are spatially local because the destruction of the Kondo resonance is a local phenomenon. The Ginzburg–Landau action will now contain, in addition to the ϕ^4 component S_{lw} , parts describing these new local critical degrees of freedom and their nonlinear coupling to the long-wavelength modes. The explicit construction of these additional parts of the Ginzburg–Landau action is left for future work.

Third, consideration of a Ginzburg–Landau description makes it likely that the locally critical point is stable beyond our EDMFT approximation, provided that $\alpha < 1$ (as is the case in $\text{CeCu}_{6-x}\text{Au}_x$ and YbRh_2Si_2). The crucial question is what happens when we allow an explicit $\mathbf{q}$ dependence in the spin self-energy. The dynamical exponent entering S_{lw} will be $z = 2/\alpha$. When $\alpha < 1$, the effective dimension $D_{\text{eff}} = D + z = 2 + 2/\alpha$ will be above the upper critical dimension of 4, so that all the nonlinear couplings within S_{lw} will be irrelevant in the renormalization-group sense¹. As a result, the

$\mathbf{q}$ -dependence of the spin self-energy from S_{lw} will be at most $(\mathbf{q} - \mathbf{Q})^2$. We also expect that the spin self-energy due to the coupling of the order-parameter field with the local modes will have a smooth $\mathbf{q}$ -dependence. As a result, the zero-temperature spin susceptibility should be proportional to $(\mathbf{q} - \mathbf{Q})^{-2}$ (that is, the anomalous exponent $\eta = 0$). This implies that the singular local susceptibility and the concomitant destruction of the Kondo resonance at the critical point are robust in two dimensions. We note that a rigorous proof of this stability awaits the explicit construction of the Ginzburg–Landau action for the locally critical point.

Fourth, Ginzburg–Landau considerations also provide a way to understand the ω/T scaling (equation (2)) at the locally critical point. At a spin-density-wave transition, the field theory, equation (15), is a ϕ^4 theory with an effective dimensionality $D_{\text{eff}} = D + z$. Whereas at QCPs in insulators^{1,36} the dynamical exponent is usually 1, in a metallic environment $z = 2$, raising D_{eff} above the upper critical dimension of 4 for the ϕ^4 theory. All the nonlinear couplings are then irrelevant, and ω/T scaling is absent^{1,14}. (A violation of ω/T scaling has been reported³⁷ near a QCP in $\text{Ce}_{1-x}\text{La}_x\text{Ru}_2\text{Si}_2$.)

At a locally critical point, the nonlinear couplings among the long-wavelength modes remain irrelevant. However, the Ginzburg–Landau action also contains nonlinear couplings that involve the additional local critical modes. These couplings are relevant, making the field theory an interacting one, thereby³⁸ allowing the ω/T scaling given in equation (2). In terms of a suitably defined spin relaxation rate, the contribution from the ϕ^4 component depends on temperature with a power greater than 1. On the other hand, the contribution from the coupling to the local modes is linear in temperature due to the ‘relevant’ nonlinear couplings. The total relaxation rate is then linear in temperature, as required³⁹ for ω/T scaling. Again, the EDMFT analysis of the previous section amounts to a microscopic prescription for determining the relaxation rate and, more generally, the entire scaling function.

Broader implications

In addition to explaining the salient features of experiments on heavy-fermion metals close to a QCP, our results may have broader implications for other strongly correlated metals. The emergence of critical local modes in the case that we have studied depends crucially on the formation of local moments. A local moment is nothing other than a strongly interacting f -electron orbital that cannot be doubly occupied owing to a strong on-site Coulomb interaction²¹. It is formed at energies intermediate between the bare Coulomb interaction scale and the asymptotic low-energy limit. The existence of non-trivial local physics at such intermediate energy scales is in fact ubiquitous in strongly correlated metals, including metals close to being a Mott insulator. This suggests the possibility of a locally critical point occurring in a doped Mott insulator. $\square$

Methods

We adopt the EDMFT as a conserving resummation of diagrams for finite dimensional systems (Fig. 7b of ref. 24). The spatial dimensionality enters, among other ways, via the form of the RKKY density of states defined in equation (8). For any finite-dimensional system, the support to $\rho_f(\epsilon)$ is bounded and a stable paramagnetic solution exists.

The EDMFT provides a self-consistent procedure. In the Bose–Fermi Kondo Hamiltonian, equation (6), we insert trial forms for the spectra of both the vector-bosonic bath and fermionic bath, which can be parametrized in terms of the Weiss fields χ_0^{-1} and G_0^{-1} , respectively:

$$g^2 \sum_p \frac{2w_p}{\omega^2 - w_p^2} = -\chi_0^{-1}(\omega), \quad \sum_p \frac{1}{\omega - E_p} = G_0(\omega) \quad (16)$$

The trial forms for the QCP are taken to be

$$\begin{aligned} \text{Im } G_0(\omega + i0^+) &= -\pi N_0, \\ \text{Im } \chi_0^{-1}(\omega + i0^+) &= C|\omega|^\gamma \text{sgn } \omega \text{ for } |\omega| < \Lambda \end{aligned} \quad (17)$$

where N_0 , γ , C , and Λ are parameters to be determined and ‘sgn’ refers to the sign function.

Next, we solve the Bose–Fermi Kondo problem using a $(1 - \gamma)$ expansion. This step

determines the local spin susceptibility and the local conduction electron Green's function,

$$\begin{aligned}\chi_{\text{loc}}(\tau) &= -\langle T_\tau S_x(\tau) S_x(0) \rangle_{H_{\text{loc}}} \\ G_{\text{loc}}(\tau) &= -\langle T_\tau c_\sigma(\tau) c_\sigma^\dagger(0) \rangle_{H_{\text{loc}}}\end{aligned}\quad (18)$$

where τ is the imaginary time. This step also specifies²⁴ the spin self-energy $M(\omega)$, and the conduction electron self-energy $\Sigma(\omega)$:

$$\begin{aligned}M(\omega) &= \chi_0^{-1}(\omega) + \frac{1}{\chi_{\text{loc}}(\omega)} \\ \Sigma(\omega) &= G_0^{-1}(\omega) - \frac{1}{G_{\text{loc}}(\omega)}\end{aligned}\quad (19)$$

(The spin self-energy is defined in terms of an effective spin cumulant that is I -irreducible²⁴.)

The $(1 - \gamma)$ expansion we use follows refs 28, 29, 34. To linear order in $1 - \gamma$, there is a renormalization-group fixed point located at $J_K = C = (1 - \gamma)/2$. In analogy to the standard application of the ϵ expansion⁴⁰, we assume that the local spin susceptibility $\chi_{\text{loc}}(\tau)$ takes a power-law form at the critical point. By setting J_K and C to their fixed point values, we can determine the exponent by matching the leading $\ln(\tau)$ term in the expansion of the assumed power-law form for $\chi_{\text{loc}}(\tau)$ with the leading $\ln(\tau)$ term of the perturbative expression for $\chi_{\text{loc}}(\tau)$, calculated from the Bose–Fermi Kondo hamiltonian, equation (6), to linear order in $1 - \gamma$. This gives, for $0 < \gamma < 1$,

$$\chi_{\text{loc}}(\omega) = \frac{1}{2\Lambda^{1-\gamma}} \Gamma(\gamma) \sin \frac{\pi(1-\gamma)}{2} (-i\omega)^{-\gamma} \quad (20)$$

where Γ is the gamma function, and for $\gamma = 0$, equation (10).

In addition, to linear order in $1 - \gamma$, the conduction electron self-energy vanishes:

$$\Sigma = O((1 - \gamma)^2) \quad (21)$$

Finally, we demand self-consistency, which amounts to the requirement that a local correlation function is equal to the wavevector average of the corresponding lattice correlation function. In terms of the RKKY density of states, $\rho_l(\epsilon)$ defined in equation (8), and the usual conduction electron density of states, $\rho_0(\epsilon)$, the self-consistency condition can be written as

$$\begin{aligned}\chi_{\text{loc}}(\omega) &= \int d\epsilon \frac{\rho_l(\epsilon)}{M(\omega) + \epsilon} \\ G_{\text{loc}}(\omega) &= \int d\epsilon \frac{\rho_0(\epsilon)}{\omega - \epsilon - \Sigma(\omega)}\end{aligned}\quad (22)$$

When $\rho_l(\epsilon)$ is proportional to $\sqrt{\epsilon - I_Q}$ near its lower band edge ($\epsilon \rightarrow I_Q^+$), χ_{loc} is finite at the QCP. The local moments are not critical.

The locally critical point arises when $\rho_l(\epsilon)$ has a jump at the lower band edge. χ_{loc} is now singular at the QCP. The self-consistent solution for the parameters introduced in equations (17) is

$$N_0 = \rho_0(\mu), \quad \gamma = 0^+, \quad C = \pi\Lambda \quad (23)$$

with the expression for Λ being given in the main text, as are those for χ_{loc} and M . We also note that the relationship between γ and α is indirect: it is specified by the self-consistency equation (22) along with equations (17)–(19).

The local susceptibility, given in equation (10), is universal. The exponent α for the spin self-energy depends on the product $\Lambda\rho_l(I_Q)$ (see equation (13)). It can be seen from equation (11) that Λ is of the order of the Kondo energy. As the QCP is reached through a competition between the RKKY and Kondo interactions, we expect that this product is close to unity, resulting in an α that is not too far from $1/2$. The precise value of α , however, is interaction-dependent: it depends on which point of the phase boundary in the RKKY interaction–Kondo energy parameter space is crossed as the system is tuned through the quantum phase transition.

At finite temperatures, the self-consistent procedure outlined above can also be carried through, leading to the spin self-energy quoted in the main text.

Received 8 December 2000; accepted 29 August 2001.

1. Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, Cambridge, 1999).
2. Maple, M. B. *et al.* Non-Fermi-liquid behavior in strongly correlated f -electron materials. *J. Low Temp. Phys.* **95**, 225–243 (1994).
3. Mathur, N. D. *et al.* Magnetically mediated superconductivity in heavy fermion compounds. *Nature* **394**, 39–43 (1998).

4. von Löhneysen, H. *et al.* Non-Fermi-liquid behavior in a heavy-fermion alloy at a magnetic instability. *Phys. Rev. Lett.* **72**, 3262–3265 (1994).
5. Trovarelli, O. *et al.* YbRh₂Si₂: Pronounced non-Fermi-liquid effects above a low-lying magnetic phase transition. *Phys. Rev. Lett.* **85**, 626–629 (2000).
6. Heuser, K. *et al.* Inducement of non-Fermi-liquid behavior with a magnetic field. *Phys. Rev. B* **57**, R4198–R4201 (1998).
7. Estrela, P., de Visser, A., Naka, T., de Boer, F. R. & Pereira, L. C. J. High-pressure study of the non-Fermi liquid material U₂Pt₂In. Preprint cond-mat/0009324 at (<http://xxx.lanl.gov>) (2000).
8. Steward, G. Non-Fermi liquid behavior in d- and f- electron metals. *Rev. Mod. Phys.* **73**, in the press (2001).
9. Schröder, A., Aeppli, G., Bucher, E., Ramazashvili, R. & Coleman, P. Scaling of magnetic fluctuations near a quantum phase transition. *Phys. Rev. Lett.* **80**, 5623–5626 (1998).
10. Stockert, O., von Löhneysen, H., Rosch, A., Pyka, N. & Loewenhaupt, M. Two dimensional fluctuations at the quantum critical point of CeCu_{6-x}Au_x. *Phys. Rev. Lett.* **80**, 5627–5630 (1998).
11. Schröder, A. *et al.* Onset of antiferromagnetism in heavy-fermion metals. *Nature* **407**, 351–355 (2000).
12. Lonzarich, G. G. in *Electron* (ed. Springford, M.) 109–147 (Cambridge Univ. Press, Cambridge, 1997).
13. Hertz, J. A. Quantum critical phenomena. *Phys. Rev. B* **14**, 1165–1184 (1976).
14. Millis, A. J. Effect of a nonzero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48**, 7183–7196 (1993).
15. Coleman, P. Theories of non-Fermi liquid behavior in heavy fermions. *Physica B* **259–261**, 353–358 (1999).
16. Si, Q., Smith, J. L. & Ingersent, K. Quantum critical behavior in Kondo systems. *Int. J. Mod. Phys. B* **13**, 2331–2342 (1999).
17. Rosch, A. Interplay of disorder and spin fluctuations in the resistivity near a quantum critical point. *Phys. Rev. Lett.* **82**, 4280–4283 (1999).
18. Coleman, P., Pépin, C. & Tsvelik, A. M. Supersymmetric spin operators. *Phys. Rev. B* **62**, 3852–3868 (2000).
19. Leggett, A. J. *et al.* Dynamics of the dissipative two-state system. *Rev. Mod. Phys.* **59**, 1–86 (1987).
20. Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, Cambridge, 1993).
21. Anderson, P. W. Localized magnetic states in metals. *Phys. Rev.* **124**, 41–53 (1961).
22. Doniach, S. The Kondo lattice and weak antiferromagnetism. *Physica B* **91**, 231–234 (1977).
23. Varma, C. M. Mixed-valence compounds. *Rev. Mod. Phys.* **48**, 219–238 (1976).
24. Smith, J. L. & Si, Q. Spatial correlations in dynamical mean-field theory. *Phys. Rev. B* **61**, 5184–5192 (2000).
25. Si, Q. & Smith, J. L. Kosterlitz–Thouless transition and short range spatial correlations in an extended Hubbard model. *Phys. Rev. Lett.* **77**, 3391–3394 (1996).
26. Chitra, R. & Kotliar, G. Effect of Coulomb long-range interactions on the Mott transition. *Phys. Rev. Lett.* **84**, 3678–3681 (2000).
27. Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13–125 (1996).
28. Smith, J. L. & Si, Q. Non-Fermi liquids in the two-band extended Hubbard model. *Europhys. Lett.* **45**, 228–234 (1999).
29. Sengupta, A. M. Spin in a fluctuating field: the Bose (+Fermi) Kondo models. *Phys. Rev. B* **61**, 4041–4043 (2000).
30. Sachdev, S. & Ye, J. Gapless spin-fluid ground state in a random quantum Heisenberg magnet. *Phys. Rev. Lett.* **70**, 3339–3342 (1993).
31. Continentino, M. A. Universal behavior in heavy fermions. *Phys. Rev. B* **47**, 11587–11590 (1993).
32. Rosch, A., Schröder, A., Stockert, O. & von Löhneysen, H. Mechanism for the non-Fermi-liquid behavior in CeCu_{6-x}Au_x. *Phys. Rev. Lett.* **79**, 159–162 (1997).
33. Aronson, M. C. *et al.* Non-Fermi-liquid scaling of the magnetic response in UCu_{5-x}Pd_x ($x = 1, 1.5$). *Phys. Rev. Lett.* **75**, 725–728 (1995).
34. Vojta, M., Buragohain, C. & Sachdev, S. Quantum impurity dynamics in two-dimensional anti-ferromagnets and superconductors. *Phys. Rev. B* **61**, 15152–15184 (2000).
35. Abanov, A. & Chubukov, A. V. Spin-fermion model near the quantum critical point: one-loop renormalization group results. *Phys. Rev. Lett.* **84**, 5608–5611 (2000).
36. Chakravarty, S., Halperin, B. I. & Nelson, D. R. Two-dimensional quantum Heisenberg antiferromagnet at low temperatures. *Phys. Rev. B* **39**, 2344–2371 (1989).
37. Raymond, S., Regnault, L. P., Flouquet, J., Wildes, A. & Lejay, P. Pressure dependence of the spin dynamics around a quantum critical point: an inelastic neutron scattering study of Ce_{0.87}La_{0.13}Ru₂Si₂. Preprint cond-mat/0102427 at (<http://xxx.lanl.gov>) (2001).
38. Sachdev, S. Theory of finite-temperature crossovers near quantum critical points close to, or above, their upper-critical dimensions. *Phys. Rev. B* **55**, 142–163 (1997).
39. Varma, C. M., Littlewood, P. B., Schmitt-Rink, S., Abrahams, E. & Ruckenstein, A. E. Phenomenology of the normal state of Cu–O high temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).
40. Wilson, K. G. & Kogut, J. The renormalization group and the ϵ expansion. *Phys. Rep. C* **12**, 75–200 (1974).

Acknowledgements

We thank G. Aeppli, A. Chubukov, P. Coleman, A. J. Millis, A. Schröder, A. M. Sengupta, C. M. Varma and P. Wölfle for discussions. This work was supported by the NSF, TCSUH and the A. P. Sloan Foundation.

Correspondence and requests for materials should be addressed to Q.S. (e-mail: qmsi@rice.edu).

Ion conduction pore is conserved among potassium channels

Zhe Lu, Angela M. Klem & Yajamana Ramu

Department of Physiology, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia, Pennsylvania 19104, USA

Potassium channels, a group of specialized membrane proteins, enable K^+ ions to flow selectively across cell membranes. Transmembrane K^+ currents underlie electrical signalling in neurons and other excitable cells. The atomic structure of a bacterial K^+ channel pore has been solved by means of X-ray crystallography. To the extent that the prokaryotic pore is representative of other K^+ channels, this landmark achievement has profound implications for our general understanding of K^+ channels. But serious doubts have been raised concerning whether the prokaryotic K^+ channel pore does actually represent those of eukaryotes. Here we have addressed this fundamental issue by substituting the prokaryotic pore into eukaryotic voltage-gated and inward-rectifier K^+ channels. The resulting chimaeras retain the respective functional hallmarks of the eukaryotic channels, which indicates that the ion conduction pore is indeed conserved among K^+ channels.

Potassium channels are a highly diversified group of tetrameric integral membrane proteins. Voltage-gated K^+ channels (K_v) underlie electrical impulse generation in nerve, muscle and endocrine cells, whereas inward-rectifier K^+ (Kir) channels help to maintain and regulate resting membrane potential. K_v channel subunits have six transmembrane segments, whereas Kir subunits have two^{1–5}. The atomic structure of a bacterial K^+ (KcsA) channel pore has been solved by X-ray crystallography⁶. To the extent that the pore of KcsA is representative of other classes of K^+ channels, the implication of this achievement for a general understanding of K^+ channels is obvious.

Although KcsA and Kir subunits share a two-transmembrane-helix topology, their primary structures are barely conserved^{4,5,7}, and serious doubts about the generality of the KcsA pore among K^+ channels have been raised. For example, it has been suggested, on the basis of a genetic study, that the arrangement of transmembrane segments of eukaryotic Kir and of prokaryotic KcsA is distinctly different⁸. It has also been proposed that a sharp bend—presumably occurring at a conserved Pro-x-Pro sequence in the sixth transmembrane segment (S6)—makes the inner pore of K_v channels different from that of KcsA, with significant implications for channel gating⁹.

Substitution of the KcsA pore in a K_v channel

To assess the generality of the KcsA pore, we produced a chimaeric channel, KcsA–Shaker, in which the prokaryotic KcsA pore replaces that of the eukaryotic Shaker K_v channel with its inactivation particle removed¹⁰ (Fig. 1a). The current through the chimaeric channels is both K^+ selective and voltage gated (Fig. 1b–e; 10 mM K^+ in extracellular solution). Although the chimaera's gating parameters differ quantitatively from those of Shaker itself (the depolarization-induced current develops more slowly than in Shaker; $V_{1/2}$ of the activation curve is about +10 mV compared with less than –30 mV for Shaker; and its apparent valence is 1.5 compared with more than 3 for Shaker)^{11,12}, they resemble those of Shaker channels that contain certain point mutations. For example, $V_{1/2}$ and the apparent valence of the activation curve for Shaker channels containing a leucine to tryptophan mutation at residue 398 are about +20 mV and 1.8, respectively¹¹.

Extracellular agitoxin 2 (AgTx2) blocks Shaker, but not KcsA channels^{13,14} or our KcsA–Shaker construct (Fig. 2c). But by replacing Gln 58 of KcsA with alanine, and Thr 61 and Arg 64 with their counterparts in Shaker's P-loop (Ser 428 and Asp 431), KcsA is rendered sensitive to AgTx2 (ref. 14). Thus, if the attached Shaker

sequences do not significantly alter the pore structure of the chimaera, then a KcsA–Shaker construct with the three point mutations (KcsA_{3M}–Shaker) should also be sensitive to AgTx2. This is verified in Fig. 2b, d and e, which shows that KcsA_{3M}–Shaker is 20 times more sensitive than Shaker to AgTx2.

The inward current at the –100 mV holding potential in oocytes expressing KcsA–Shaker channels, which is minimal in 10 mM extracellular K^+ (Figs 1b and 3b) but prominent in 100 mM K^+ (Fig. 3d), is larger than the expected background leak current in oocytes not injected with RNA. This observation, and the fact that the inward current at –100 mV in KcsA_{3M}–Shaker chimaeras is sensitive to AgTx2 (Fig. 3f) argue that the chimaeric channels still conduct at very negative potentials—a feature resembling that of a P475D Shaker mutant whose activation gate is suggested to be disrupted¹⁵. This raises the question of whether our chimaeric construct is only partially voltage gated (that is, cannot be fully closed) or whether hyperpolarization beyond experimental reach is required to close it completely. Although this uncertainty remains, KcsA–Shaker with an alanine mutation at the position corresponding to Gly 104 in KcsA opens at more depolarized potentials and carries no significant current at –100 mV with either low or high extracellular K^+ concentration (Fig. 3c, e, g). This finding confirms that the chimaeric construct is intrinsically capable of being fully gated by membrane voltage. In addition, the current carried by KcsA–Shaker decreases (less so with the G104A mutation) as extracellular K^+ is lowered, a dependence observed previously with other K_v channels¹⁶, which may result from channel block and/or a conformational change as in the case of C-type inactivation^{17–19}.

Substitution of the KcsA pore in a Kir channel

Whereas the amino-acid sequence that forms the pore is well conserved between K_v and KcsA, it is markedly different in Kir, as illustrated for the amino-terminal portion of the P-loop, which is defined here as the entire linker between the first and second transmembrane segments (M1 and M2) in Kir and KcsA, or its counterpart between S5 and S6 in K_v channels (Fig. 4a). Alignment shows that the P-loops in Kir differ from those in K_v and KcsA not only in residue composition, but also in length on account of an extra, variable segment that presumably forms the turrets lining the external vestibule in Kir⁶. However, deletion of this 'insert' from IRK1 (Kir 2.1), which reduces the length of its P-loop to that of KcsA and K_v , has little effect on either voltage-jump-induced relaxation or the degree of rectification (Fig. 4b, c).

More direct evidence for pore conservation between KcsA and Kir

comes from a chimera, KcsA–IRK1, in which we replaced the pore of IRK1 with that of KcsA (Fig. 5a, b). Like KcsA²⁰, the chimera is blocked by extracellular tetraethylammonium (TEA) (dissociation constant, $K_d = 0.8$ mM) with little voltage dependence (Fig. 5d, f, h, i). The fact that KcsA–IRK1 rectifies only modestly compared with IRK1 is probably due to the absence in KcsA of an aspartate residue that is present in the M2 region of IRK1 and is crucial for the voltage-dependent high-affinity binding of intracellular blocking cations that underlies strong inward rectification^{21–25}.

Because replacing a neutral residue with aspartate in the M2 region of weakly rectifying ROMK1 (Kir 1.1) is known to enhance markedly that channel's affinity for intracellular cations and hence rectification^{26,27}, we made such a substitution in KcsA–IRK1. The M2 residue replaced, which corresponds to Thr 107 in KcsA, sits at the intracellular end of the cavity that separates the outer K⁺-selective part from the inner part of the pore, its side chain pointing into the cavity⁶. The T107D mutation does convert KcsA–IRK1 into a strong inward rectifier (Fig. 5c). The small outward current in the

oocyte expressing T107D channels is not reduced much by extracellular TEA at concentrations that strongly block the much larger inward current, and must therefore be background current (Fig. 5e, g). This signifies that T107D channels themselves carry insignificant outward current. Current relaxation in T107D channels after membrane hyperpolarization is much slower than in IRK1—a feature seen in some other strong rectifiers including GIRK1/4 (Kir 3.1/Kir 3.4)^{28–30} and BIR10 (Kir 4.1)³¹.

Mutation T107D in the KcsA pore not only enhances rectification in KcsA–IRK1 but also renders voltage-gated KcsA–Shaker inwardly rectifying. At a holding potential of 0 mV, KcsA_{3M}–Shaker exhibits significant (AgTx2-sensitive) conductance, which undergoes further activation with time at positive voltages and considerable deactivation at negative voltages (Fig. 6b, d). In contrast, KcsA_{3M}–Shaker containing T107D carries large AgTx2-sensitive inward currents but minimal outward currents; no deactivation is observed at negative voltages (Fig. 6c). This point mutation therefore converts a delayed rectifier into an inward rectifier (Fig. 6b, d, compare c, e), although the extent of intrinsic voltage gating of the T107D-containing channels, if any, remains to be determined. Thr 107 corresponds to Val 474 of Shaker in the conserved Pro-x-Pro sequence of classic K_v channels⁹ (Fig. 6a). Evidently, when examined under comparable conditions, KcsA pores with the T107D mutation conduct K⁺ in a comparable manner whether they are inserted into an IRK1 or a Shaker sequence (compare Figs 5 and 6).

Discussion

Our findings show that the pore of KcsA can substitute for that of K_v or Kir. Therefore, although they differ fundamentally in both overall architecture (two compared with six transmembrane segments) and gating properties (K_v is gated by voltage, KcsA by protons^{20,32}, and Kir by PIP₂ (ref. 33) and other messenger molecules), these K⁺ channels must share a conserved pore. Our argument for a conserved pore is consistent with the previously reported promis-

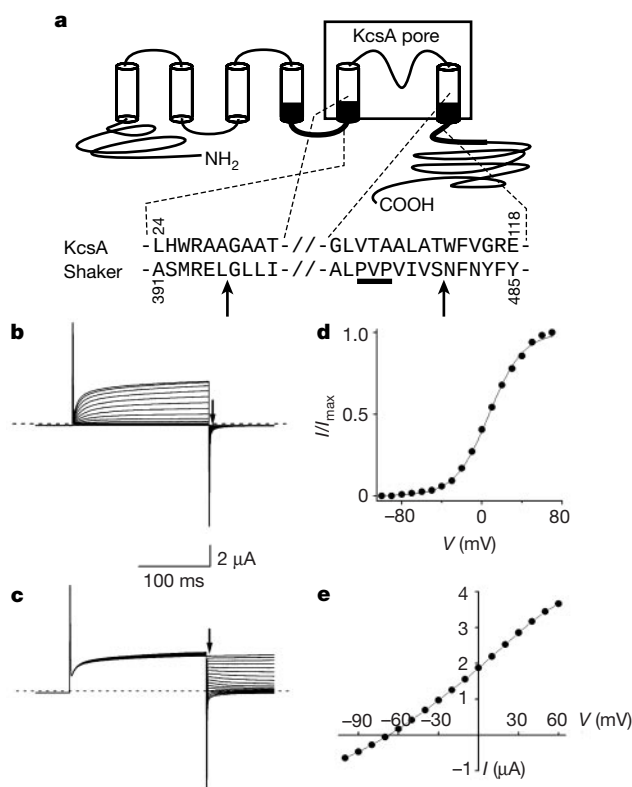


Figure 1 Replacement of the Shaker pore module with that of KcsA. **a**, Representation of the polypeptide chain topology of the KcsA–Shaker chimeric channel and partial sequences of the parent channels around the splicing sites. In the chimera, the S5–S6 region of Shaker (indicated between the arrows) is replaced by the corresponding part from KcsA. Black regions, putative interaction regions (see Discussion). **b**, **c**, KcsA–Shaker currents recorded in 10 mM K⁺ solution from the same oocyte, as the membrane potential was stepped from the –100 mV holding potential either to test potentials 10 mV apart between –100 and +70 mV, and back (**b**), or to +60 mV and then test potentials between –100 and +60 mV (**c**). Dashed line identifies the zero current level.

d, **e**, Currents in **b** (normalized) and **c**, respectively, at the time indicated by arrows are plotted against voltage. As an approximation, data in **d**, where current at –100 mV is set as zero and that at +70 mV as $I_{\max}$, are fitted with the equation

$$\frac{I}{I_{\max}} = \frac{1}{1 + \exp\left(\frac{-zF(V - V_{1/2})}{RT}\right)}$$

(where F , R and T are Faraday's constant, the gas constant and absolute temperature, respectively) yielding valence $z = 1.5$ and midpoint voltage $V_{1/2} = +7$ mV.

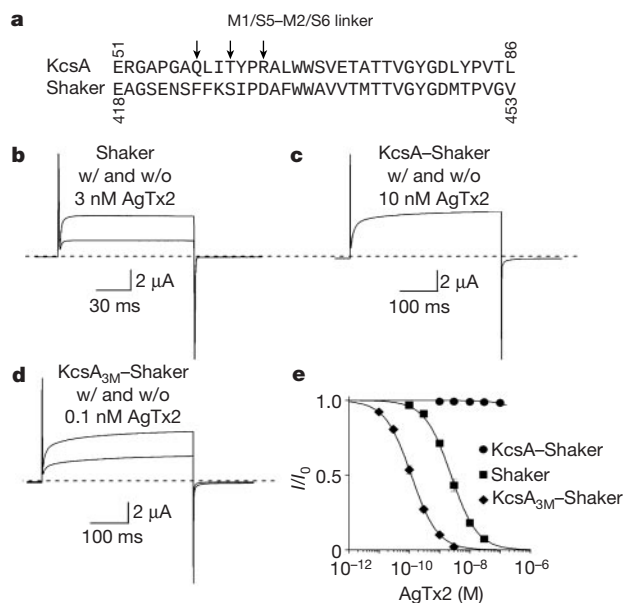


Figure 2 Triple mutation in the P-loop renders KcsA–Shaker channels sensitive to AgTx2.

a, Alignment of the KcsA and Shaker P-loops. Arrows indicate the sites of the three point mutations described in the text. **b–d**, Currents through Shaker, KcsA–Shaker or KcsA_{3M}–Shaker channels with and without AgTx2, recorded as the membrane potential was stepped from –80 mV to +20 mV and back. **e**, Fractions of unblocked currents (mean $\pm$ s.e.m., $n = 6–7$) versus AgTx2 concentration, where I_0 is the current in the absence of AgTx2. The curves through the data of Shaker and KcsA_{3M}–Shaker are fits of the equation $I/I_0 = K_d / (K_d + [AgTx2])$, with $K_d = 2.38 \pm 0.12$ nM (mean $\pm$ s.e.m., $n = 6$) and 0.11 ± 0.01 nM ($n = 7$), respectively.

cuity displayed by the external pore-blocking scorpion toxins among various classes of K^+ channels, including K_v , Kir and $KcsA$ ^{14,34}. Some differences among K^+ channel pores must undoubtedly remain. For example, the pore of IRK1 may be extended intracellularly by part of the carboxy terminus beyond the essential, conserved pore-forming module^{25,35}. In addition, a conserved

Pro-x-Pro sequence in the S6 region of classic K_v channels may confer specific features to those channels⁹, although it is not essential for voltage gating because $KcsA$ -Shaker, which lacks this sequence, is still voltage gated (Fig. 3).

Our chimaeric channel constructs illustrate experimentally the modular nature of K_v channel architecture, namely, a pore-forming module plus one or more voltage-sensing modules^{4,5,7,11,36}. This modular architecture implies that voltage-gated K^+ channels evolved from $KcsA$ or similar channels and that a separate $KcsA$ (voltage or non-voltage) gating module, not attached to the pore

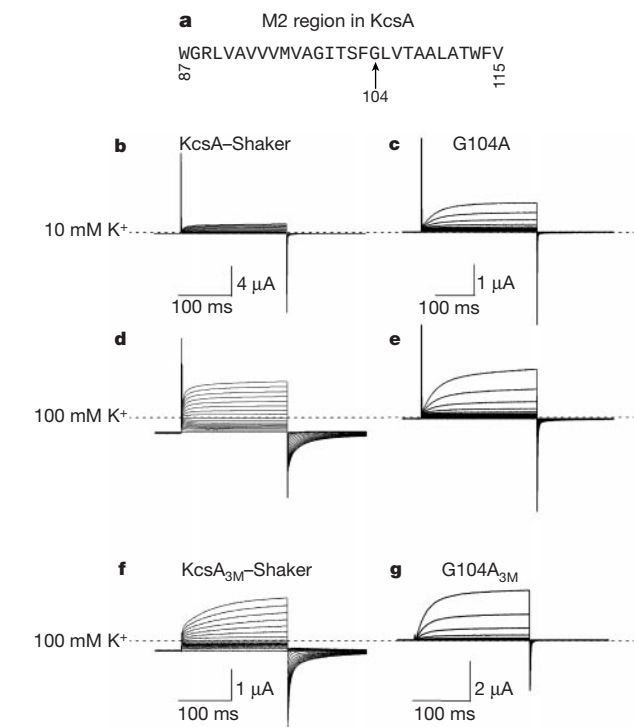


Figure 3 Effects of extracellular K^+ concentration and a G104A mutation on the gating of $KcsA$ -Shaker. **a**, Amino-acid sequence of the M2 region of $KcsA$. **b-e**, Currents through $KcsA$ -Shaker without and with an alanine mutation at the residue corresponding to G104 of $KcsA$ in the absence or presence of two concentrations of extracellular K^+ . Records shown in **b** and **d**, or in **c** and **e** were recorded from the same oocyte. **f, g**, Currents recorded through $KcsA_{3M}$ -Shaker channels without and with the G104A mutation are corrected for background current with AgTx2 as described in Methods. The voltage protocol was the same as used in Fig. 1b.

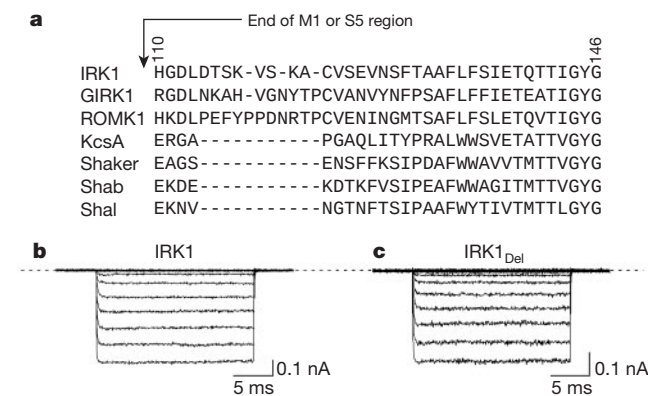


Figure 4 Deletion of a segment in the P-loop of the IRK1 channel. **a**, Alignment of the partial P-loop from N-terminal end⁴ to Gly-Tyr-Gly among various K^+ channels^{2-5,7,28,29,50}. **b, c**, Currents of IRK1 and of its mutant lacking the sequence between His-Gly-Asp-Leu and Cys-Val-Ser-Glu (IRK1_{Del}), corrected for background leak currents as described in Methods. To resolve the relatively rapid current relaxation kinetics in these channels, we recorded from cell-attached membrane patches rather than from whole oocytes (which were used in the other figures). During the recording, membrane potential was stepped from the 0 mV holding potential to test potentials 10 mV apart between -80 and +80 mV, and then back to 0 mV.

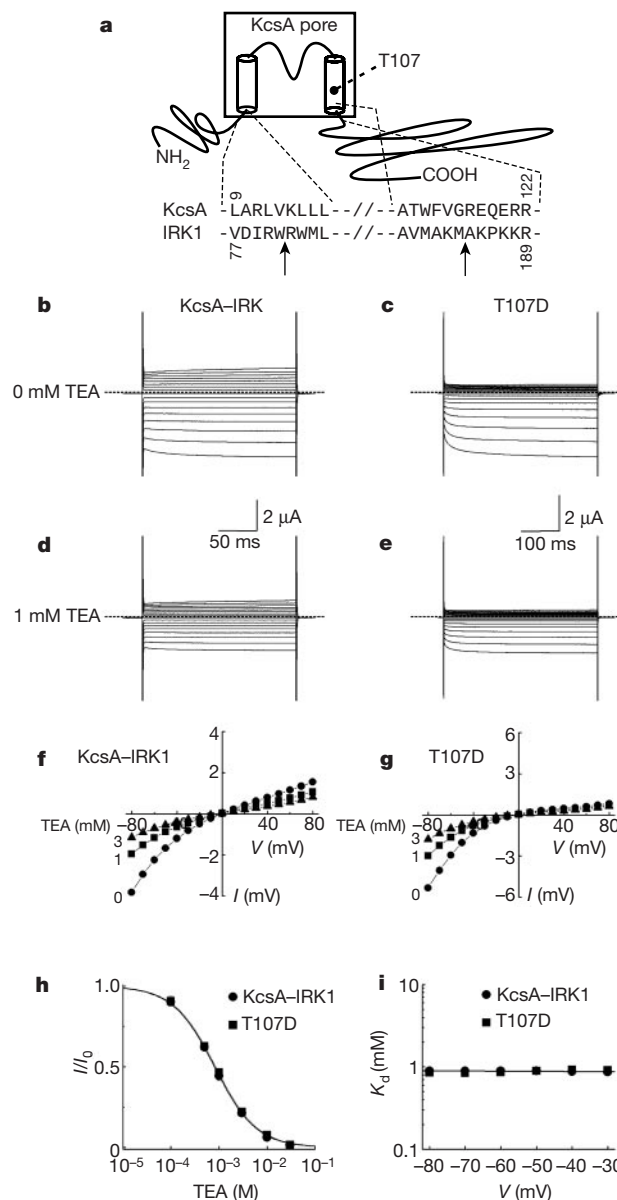


Figure 5 Substitution of the $KcsA$ pore in IRK1. **a**, Peptide chain topology of the $KcsA$ -IRK1 chimaeric channel and partial sequences of the parent channels around the splicing sites. In the chimaera, the M1-M2 region of IRK1 (indicated between the arrows) is replaced by the corresponding part from $KcsA$. **b-g**, Currents through $KcsA$ -IRK1 (**b, d**) and its T107D mutant (**c, e**), without (**b, c**) or with (**d, e**) extracellular TEA, were recorded using the voltage protocol of Fig. 4; the corresponding I - V curves are plotted in **f** and **g**. **h**, Fractions of unblocked currents (mean $\pm$ s.e.m., $n = 5$) at -80 mV are plotted against TEA concentration; the curves are fits of the equation $I/I_0 = K_d / (K_d + [TEA])$ where $K_d = 0.81 \pm 0.02$ mM for $KcsA$ -IRK1 and 0.86 ± 0.02 mM for T107D. **i**, K_d (mean $\pm$ s.e.m., $n = 5$) determined from fits similar to (and including) those shown for -80 mV in **h** are plotted against membrane voltage.

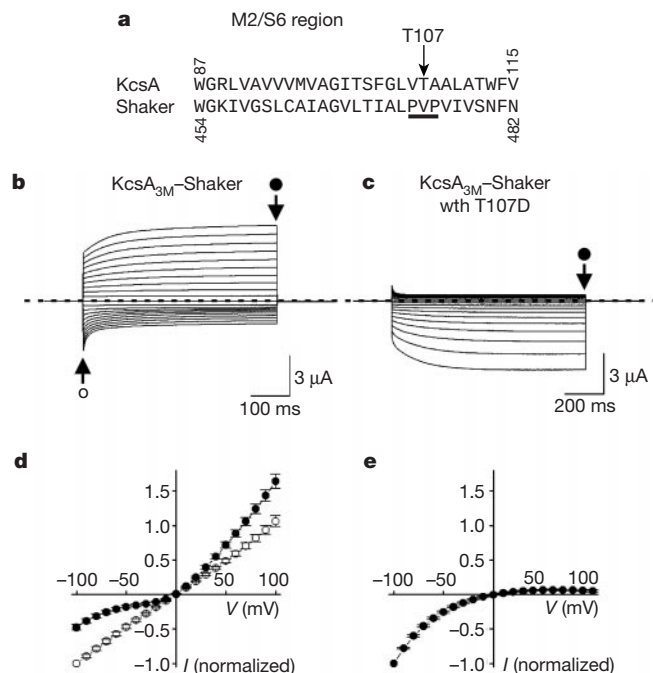


Figure 6 Mutation T107D renders KcsA–Shaker inwardly rectifying. **a**, Alignment of M2/S6 sequences of KcsA and Shaker. **b**, **c**, Current records of KcsA_{3M}–Shaker without and with T107D in the presence of 100 mM extracellular K⁺. During the recording, membrane potential was stepped from the 0 mV holding potential to test potentials 10 mV apart between –100 and +100 mV, and then back to 0 mV. Correction for background currents was done with AgTx2 as described in Methods. **d**, **e**, *I*–*V* curves constructed from data similar to (and including) those shown in **b** and **c**, at the time indicated by arrows and with the corresponding symbols. Each data point represents mean current ($\pm$ s.e.m., $n = 13$ (**d**) and $n = 12$ (**e**)) normalized to that at –100 mV. The AgTx2-insensitive outward current in oocytes injected with cDNA of KcsA_{3M}–Shaker containing T107D is significantly larger than in uninjected oocytes, and unaffected by replacing 100 mM extracellular K⁺ with 100 mM Na⁺. This current may reflect the oocyte's response to expression of the channels.

module through the main peptide chain, may exist in *Streptomyces lividans*. Furthermore, our demonstration that attaching the voltage-sensing module of Shaker to the pore of KcsA confers voltage gating on the latter has important implications regarding the mechanics of voltage gating.

For example, the voltage-sensing module has been suggested to couple the voltage sensor to the channel gate through a specific set of residues in the shell of the pore-forming module of K_v (ref. 11). If so, the voltage sensitivity of our KcsA–Shaker chimera would argue that the shell of KcsA's pore (which is presumed normally to form an interface with membrane lipids) must be constructed in such a way that it, too, can establish a functional interface that mediates coupling between the voltage-sensing module(s) and the channel gate. Considering that our voltage-gated KcsA–Shaker chimera contains KcsA's pore module (S5–S6) but Shaker's S4, S4–S5 linker and C terminus, primary coupling between the voltage-sensing S4 (refs 37–42) and the internal channel gate probably occurs more intracellularly. In this coupling scheme, residues in the sequence encompassing the C-terminal part of S4 to the N-terminal part of S5 interact with specific residues in the C-terminal part of S6 and its extension where the intracellular gate is located^{43,44} (the two putative interacting regions are highlighted in Fig. 1a).

Consistent with this view, a study has shown that a KcsA–Shaker chimera that differs from ours mainly in that it contains KcsA's C terminus, whereas ours has Shaker's, does not conduct any current over wide ranges of voltage and pH even though it reaches the

membrane⁴⁵. Furthermore, most of the distal portion of the C terminus in a K_v channel is non-essential for voltage gating⁴⁶. If the C-terminal part of S6 and its extension (or their counterparts in other K⁺ channels) indeed act as the 'door handle', then this would explain why Kir, KcsA and Ca²⁺-activated K⁺ channels are gated by intracellular molecules, whereas K⁺ channels with inside-out membrane topology⁴⁷, such as GluR0, are gated extracellularly.

Methods

Molecular biology and channel expression

Complementary DNAs for Shaker H4 with N-type inactivation removed^{43,10} and IRK1 (ref. 5) were cloned in pBluescript and/or pGEM–HES plasmids, respectively. To replace the DNA sequence encoding the pore of Shaker or IRK1 with that of KcsA, we subcloned a polymerase chain reaction (PCR)-generated DNA fragment containing the KcsA segment of interest into the recipient version of Shaker or IRK1 cDNA. The N- and C-terminal junctions in the KcsA–Shaker construct are respectively formed by joining Leu 396 of Shaker to Gly 30 of KcsA (ref. 7; residue 397 of Shaker is also glycine) and Thr 112 of KcsA to Asn 480 of Shaker; those in the KcsA–IRK1 construct are respectively formed by joining Trp 81 of IRK1 to Lys 14 of KcsA and Gly 116 of KcsA to Ala 184 of IRK1. The cDNAs encoding the remaining mutant channels were also produced using PCR mutagenesis. We confirmed all PCR-generated sequences by DNA sequencing. All cRNA were synthesized with T7 polymerase using linearized cDNA as templates. Oocytes were prepared and injected with cRNAs as described⁴⁸.

Recordings of channel currents

All channel currents were recorded from whole oocytes using a two-electrode voltage-clamp amplifier (Warner OC-725C), except those shown in Fig. 4, which were recorded from cell-attached membrane patches using a patch-clamp amplifier (Axopatch 200B) to resolve the relatively rapid current relaxation kinetics of the channels. The resistance of electrodes filled with 3 M KCl was 0.4 MΩ. Unless specified otherwise, the bath solution for studying Shaker and KcsA–Shaker contained (in mM): 90 Na⁺ (Cl[–] + OH[–]), 10 KCl, 0.3 CaCl₂, 1 MgCl₂ and 10 HEPES; pH 7.6 with NaOH. When the K⁺ concentration was increased, the Na⁺ concentration was decreased so that their sum remained 100 mM. In the case of KcsA_{3M}–Shaker (with and without G104A or T107D) the background currents were obtained in the presence of AgTx2 at concentrations greater than 100 × K_d. The concentration of AgTx2 stock was calculated using a published extinction coefficient (7.48 mM^{–1} cm^{–1} at 235 nm)¹³. The solution for studying IRK1 and IRK1–KcsA contained (in mM): 100 K⁺ (Cl[–] + OH[–]), 0.3 CaCl₂, 1 MgCl₂ and 10 HEPES; pH 7.6 with KOH. We recorded the currents shown in Fig. 4 with a patch-clamp amplifier in the cell-attached configuration. After recording, the membrane patch was excised and its inside was exposed to a solution containing 3 mM Mg²⁺ to induce a rapid channel rundown²⁶. The currents recorded after channel rundown were used as templates for the subsequent off-line background current correction. The pipette and bath solutions contained (in mM): 5 K₂EDTA, 10 K₂HPO₄ + KH₂PO₄ in a ratio yielding pH 7.6, and sufficient KCl to bring total K⁺ concentration to 100 mM (ref. 49).

Received 1 June; accepted 29 August 2001.

- Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, Y. N. & Jan, L. Y. Sequence of a probable potassium channel component encoded at Shaker locus of *Drosophila*. *Science* **237**, 770–775 (1987).
- Pongs, O. *et al.* Shaker encodes a family of putative potassium channel proteins in the nervous system of *Drosophila*. *EMBO J.* **7**, 1087–1096 (1988).
- Kamb, A., Tseng-Crank, J. & Tanouye, M. A. Multiple products of the *Drosophila* Shaker gene may contribute to potassium channel diversity. *Neuron* **1**, 421–430 (1988).
- Ho, K. *et al.* Cloning and expression of an inwardly rectifying ATP-regulated potassium channel. *Nature* **362**, 31–38 (1993).
- Kubo, Y., Baldwin, T. J., Jan, Y. N. & Jan, L. Y. Primary structure and functional expression of a mouse inward rectifier potassium channel. *Nature* **362**, 127–133 (1993).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Schrempf, H. *et al.* A prokaryotic potassium ion channel with two predicted transmembrane segments from *Streptomyces lividans*. *EMBO J.* **14**, 5170–5178 (1995).
- Minor, D. L. J., Masseling, S. J., Jan, Y. N. & Jan, L. Y. Transmembrane structure of an inwardly rectifying potassium channel. *Cell* **96**, 879–891 (1999).
- del Camino, D., Holmgren, M., Liu, Y. & Yellen, G. Blocker protection in the pore of a voltage-gated K⁺ channel and its structural implications. *Nature* **403**, 321–325 (2000).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533–538 (1990).
- Li-Smerin, Y., Hackos, D. H. & Swartz, K. J. A localized interaction surface for voltage-sensing domains on the pore domain of a K⁺ channel. *Neuron* **25**, 411–423 (2000).
- Schoppa, N. E., McCormack, K., Tanouye, M. A. & Sigworth, F. J. The size of gating charge in wild-type and mutant Shaker potassium channels. *Science* **255**, 1712–1715 (1992).
- Garcia, M. L., Garcia-Calvo, M., Hidalgo, P., Lee, A. & MacKinnon, R. Purification and characterization of three inhibitors of voltage-dependent K⁺ channels from *Leiurus quinquestriatus* var. *hebraeus* venom. *Biochemistry* **33**, 6834–6839 (1994).
- MacKinnon, R., Cohen, S. L., Kuo, A., Lee, A. & Chait, B. T. Structural conservation in prokaryotic and eukaryotic potassium channels. *Science* **280**, 106–109 (1998).
- Hackos, D. H. & Swartz, K. J. Disrupting the cytoplasmic gate in the Shaker K⁺ channel with mutations in S6. *Biophys. J.* **80**, 182a–183a (2001).
- Pardo, L. A. *et al.* Extracellular K⁺ specifically modulates a rat brain K⁺ channel. *Proc. Natl Acad. Sci. USA* **89**, 2466–2470 (1992).

17. Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Two types of inactivation in Shaker K⁺ channels: effects of alterations in the carboxy-terminal region. *Neuron* **7**, 547–556 (1991).
18. Lopez-Barneo, J., Hoshi, T., Heinemann, S. H. & Aldrich, R. W. Effects of external cations and mutations in the pore region on C-type inactivation of Shaker potassium channels. *Recept. Channels* **1**, 61–71 (1993).
19. Yellen, G., Sodickson, D., Chen, T. Y. & Jurman, M. E. An engineered cysteine in the external mouth of a K⁺ channel allows inactivation to be modulated by metal binding. *Biophys. J.* **66**, 1068–1075 (1994).
20. Heginbotham, L., LeMasurier, M., Kolmakova-Partensky, L. & Miller, C. Single *Streptomyces lividans* K⁺ channels: Functional asymmetries and sidedness of proton activation. *J. Gen. Physiol.* **114**, 551–560 (1999).
21. Stanfield, P. R. *et al.* A single aspartate residue is involved in both intrinsic gating and blockage by Mg²⁺ of the inward rectifier, IRK1. *J. Physiol. (Lond.)* **478**, 1–6 (1994).
22. Lopatin, A. N., Makhina, E. N. & Nichols, C. G. Potassium channel block by cytoplasmic polyamines as the mechanism of intrinsic rectification. *Nature* **372**, 366–369 (1994).
23. Ficker, E., Tagliatela, M., Wible, B. A., Henley, C. M. & Brown, A. M. Spermine and spermidine as gating molecules for inward rectifier K⁺ channels. *Science* **266**, 1068–1072 (1994).
24. Fakler, B. *et al.* Strong voltage-dependent inward rectification of inward rectifier K⁺ channels is caused by intracellular spermine. *Cell* **80**, 149–154 (1995).
25. Yang, J., Jan, Y. N. & Jan, L. Y. Control of rectification and permeation by residues in two distinct domains in an inward rectifier K⁺ channel. *Neuron* **14**, 1047–1054 (1995).
26. Lu, Z. & MacKinnon, R. Electrostatic tuning of Mg²⁺ affinity in an inward-rectifier K⁺ channel. *Nature* **371**, 243–246 (1994).
27. Wible, B. A., Tagliatela, M., Ficker, E. & Brown, A. M. Gating of inwardly rectifying K⁺ channels localized to a single negatively charged residue. *Nature* **371**, 246–249 (1994).
28. Kubo, Y., Reuveny, E., Slesinger, P. A., Jan, Y. N. & Jan, L. Y. Primary structure and functional expression of a rat G-protein-coupled muscarinic potassium channel. *Nature* **364**, 802–806 (1993).
29. Dascal, N. *et al.* Atrial G protein-activated K⁺ channel: expression cloning and molecular properties. *Proc. Natl Acad. Sci. USA* **90**, 10235–10239 (1993).
30. Krapivinsky, G. *et al.* The G-protein-gated atrial K⁺ channel IKACH is a heteromultimer of two inwardly rectifying K⁺ channel proteins. *Nature* **374**, 135–141 (1995).
31. Bond, C. T. *et al.* Cloning and expression of a family of inward rectifier potassium channels. *Recept. Channels* **2**, 183–191 (1994).
32. Perozo, E., Cortes, D. M. & Cuello, L. G. Structural rearrangements underlying K⁺ channel activation gating. *Science* **285**, 73–78 (1999).
33. Huang, C. L., Feng, S. & Hilgemann, D. W. Direct activation of inward rectifier potassium channels by PIP₂ and its stabilization by Gβγ. *Nature* **391**, 803–806 (1998).
34. Lu, Z. & MacKinnon, R. Purification, characterization, and synthesis of an inward-rectifier K⁺ channel inhibitor from scorpion venom. *Biochemistry* **36**, 6936–6940 (1997).
35. Tagliatela, M., Ficker, E., Wible, B. A. & Brown, A. M. C-terminus determinants for Mg²⁺ and polyamine block of the inward rectifier K⁺ channel IRK1. *EMBO J.* **14**, 5532–5541 (1995).
36. Li-Smerin, Y. & Swartz, K. J. Gating modifier toxins reveal a conserved structural motif in voltage-gated Ca²⁺ and K⁺ channels. *Proc. Natl Acad. Sci. USA* **95**, 8585–8589 (1998).
37. Liman, E. R., Hess, P., Weaver, F. & Korean, G. Voltage-sensing residues in the S4 region of a mammalian K⁺ channel. *Nature* **353**, 752–756 (1991).
38. Papazian, D. M., Timpe, L. C., Jan, Y. N. & Jan, L. Y. Alteration of voltage-dependence of Shaker potassium channel by mutations in the S4 sequence. *Nature* **349**, 305–310 (1991).
39. Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K⁺ channel. *Neuron* **16**, 1169–1177 (1996).
40. Seoh, S.-A., Sigg, D., Papazian, D. M. & Benzanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K⁺ channel. *Neuron* **16**, 1159–1167 (1996).
41. Yang, N., George, A. L. J. & Horn, R. Molecular basis of charge movement in voltage-gated sodium channels. *Neuron* **16**, 113–122 (1996).
42. Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the shaker K⁺ channel S4. *Neuron* **16**, 387–397 (1996).
43. Armstrong, C. M. Interaction of tetraethylammonium ion derivatives with the potassium channels of giant axons. *J. Gen. Physiol.* **58**, 413–437 (1971).
44. Liu, Y., Holmgren, M., Jurman, M. E. & Yellen, G. Gated access to the pore of a voltage-dependent K⁺ channel. *Neuron* **19**, 175–184 (1997).
45. Capriani, M. *et al.* Structural compatibility between the putative voltage sensor of voltage-gated K⁺ channels and the prokaryotic KcsA channel. *J. Biol. Chem.* **276**, 21070–21076 (2001).
46. VanDongen, A. M. J., Frech, G. C., Drewe, J. A., Joho, R. H. & Brown, A. M. Alteration and restoration of K⁺ channel function by deletions at the N- and C-termini. *Neuron* **5**, 433–443 (1990).
47. Chen, G.-Q., Cui, C., Mayer, M. L. & Gouaux, E. Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature* **402**, 817–821 (1999).
48. Spassova, M. & Lu, Z. Coupled ion movement underlies rectification in an inward-rectifier K⁺ channel. *J. Gen. Physiol.* **112**, 211–221 (1998).
49. Guo, D. & Lu, Z. Pore block versus intrinsic gating in the mechanism of inward rectification in the strongly-rectifying IRK1 channel. *J. Gen. Physiol.* **116**, 561–568 (2000).
50. Butler, A., Wei, A., Baker, K. & Salkoff, L. A family of putative potassium channel genes in *Drosophila*. *Science* **243**, 943–947 (1989).

Acknowledgements

We thank P. De Weer for review and discussion of our manuscript; C. M. Armstrong, R. MacKinnon and K. J. Swartz for discussions and suggestions; R. MacKinnon and E. Perozo for KcsA cDNA; K. J. Swartz for Shaker-IR cDNA subcloned in the pGEM–HESS vector; L. Y. Jan for IRK1 cDNA; and J. Yang for IRK1 cDNA subcloned in the pGEM–HESS vector. This study was supported by the National Institute of General Medical Sciences. Z.L. is the recipient of an Independent Scientist Award from the National Heart, Lung and Blood Institute.

Correspondence and requests for materials should be addressed to Z.L. (e-mail: zhelu@mail.med.upenn.edu).

Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria

Frank Schlünzen^{*†}, Raz Zarivach[‡], Jörg Harms^{*†}, Anat Bashan[‡], Ante Tocilj^{*§}, Renate Albrecht[§], Ada Yonath^{*‡} & François Franceschi[§]

^{*} Max-Planck-Research, Unit for Ribosomal Structure, Notkestrasse 85, 22603 Hamburg, Germany

[‡] Department of Structural Biology, Weizmann Institute, 76100 Rehovot, Israel

[§] Max-Planck-Institut für Molekulare Genetik, Ihnestr. 73, 14195 Berlin, Germany

[†] These authors contributed equally to this work

Ribosomes, the site of protein synthesis, are a major target for natural and synthetic antibiotics. Detailed knowledge of antibiotic binding sites is central to understanding the mechanisms of drug action. Conversely, drugs are excellent tools for studying the ribosome function. To elucidate the structural basis of ribosome–antibiotic interactions, we determined the high-resolution X-ray structures of the 50S ribosomal subunit of the eubacterium *Deinococcus radiodurans*, complexed with the clinically relevant antibiotics chloramphenicol, clindamycin and the three macrolides erythromycin, clarithromycin and roxithromycin. We found that antibiotic binding sites are composed exclusively of segments of 23S ribosomal RNA at the peptidyl transferase cavity and do not involve any interaction of the drugs with ribosomal proteins. Here we report the details of antibiotic interactions with the components of their binding sites. Our results also show the importance of putative Mg⁺² ions for the binding of some drugs. This structural analysis should facilitate rational drug design.

The high-resolution crystallographic structures of the ribosomal subunits revealed many of the molecular details of ribosomes, including the decoding centre at the 30S subunit and the peptidyl transferase centre at the 50S subunit^{1–6}. The exact mechanism for the peptidyl transferase activity, however, is still unclear^{2,7–9}. The peptidyl transferase centre is the main target site for many antibiotics and substrate analogues, such as puromycin¹⁰.

The therapeutic use of antibiotics has been severely compromised by the emergence of drug resistance in many pathogenic bacteria. The molecular mechanisms by which bacteria become resistant usually involve drug efflux, drug inactivation, or alterations in the antibiotic target site. Although ribosomal proteins can affect the binding and action of ribosome-targeted antibiotics, the primary target of these antibiotics is rRNA¹¹. Many cases of antibiotic resistance in clinical strains can be linked to alterations of specific nucleotides of the 23S rRNA within the peptidyl transferase centre¹². However, to date the exact mode of interaction and the chemical moieties involved in binding have not been determined. The structural basis for this antibiotic action is also unknown.

Complexes of ribosomal subunits with antibiotics are essential to understand the molecular mechanisms of antibiotic binding,

antibiotic resistance, and for rational drug design. Conversely, antibiotics have been extremely useful to elucidate functional aspects of protein biosynthesis and to define the structural components involved in ribosome function.

To determine ribosome–antibiotic interactions in eubacteria and to understand better the mechanism of the peptidyl transferase activity, we determined the structure of the 50S subunit of *D. radiodurans* complexed with each of the following antibiotics: chloramphenicol, clindamycin (a lincosamide), and the macrolides erythromycin, clarithromycin and roxithromycin. Analysis of the obtained X-ray structures allowed us to propose the structural principles for the action of the antibiotics examined.

Results

The 3.1 Å structure model of the 50S subunit of *D. radiodurans*¹³ was used as a reference to determine the structures of the 50S–antibiotic complexes (Table 1). The electron density maps of the 50S–antibiotic complexes allowed us to unambiguously determine the binding sites of the five antibiotics. The chemical nature of the 50S–antibiotic interactions could be deduced largely from the mode of binding. On the basis of our difference electron density maps and on the available biochemical and functional data, we

Table 1 Crystallographic data

Parameter	Chloramphenicol	Clindamycin	Erythromycin*	Clarithromycin*	Roxithromycin*
Resolution (Å)	25–(3.62–3.5)	35–(3.21–3.1)	20–(3.62–3.5)	50–(3.62–3.5)	50–(3.87–3.8)
<i>R</i> _{sym} (%)	12.4 (57.7)	11.8 (57.0)	12.2 (45.8)	9.8 (50.6)	10.7 (24.4)
Completeness (%)	88.5 (71.0)	94.4 (82.2)	79.5 (71.4)	85.2 (78.1)	64.8 (66.8)
Redundancy	3.9	3.8	5.4	4.9	3.7
<i>I</i> / <i>σ</i> (<i>I</i>)	9.5 (2.1)	6.5 (2.1)	7.2 (1.8)	8.1 (2.1)	6.2 (2.2)
Unit cell (Å)	171.1 × 409.3 × 696.9	170.3 × 410.1 × 697.2	169.2 × 410.0 × 695.0	169.9 × 412.7 × 697.0	170.4 × 410.7 × 694.8
<i>R</i> / <i>R</i> _{free} (%)	27.5/32.1	26.8/30.3	26.8/30.1	27.3/32.3	20.9/27.4
Bond length (Å)†	0.0038	0.0050	0.0037	0.0037	0.0035
Bond angles (°)	0.69	0.77	0.66	0.65	0.67

Values in parentheses correspond to the high-resolution bin.

* Macrolide antibiotics.

† r.m.s. deviations for bond length and bond angles. The estimated coordinate error ranges from ±0.45 Å for clindamycin to ±0.65 Å for roxithromycin.

propose the structural basis for the modes of action of the antibiotics chloramphenicol, clindamycin, erythromycin, clarithromycin and roxithromycin.

Each of the antibiotics that we analysed targets the 50S subunit only at the peptidyl transferase cavity, although the binding site from each class of antibiotic differs from each other. Each class of antibiotic interacts exclusively with specific nucleotides assigned to a multi-branched loop of domain V of the 23S rRNA in the two-dimensional structure (Figs 1a, 2a and 3a), explaining previous mutational and footprinting data^{10,14,15}. None of the antibiotics show any direct interaction with ribosomal proteins. Binding of these antibiotics did not result in any significant conformational change of the peptidyl transferase cavity.

Chloramphenicol

Chloramphenicol is known to block peptidyl transferase activity by hampering the binding of transfer RNA to the A site^{16,17}. Chloramphenicol has several reactive groups that can form hydrogen bonds with various nucleotides of the peptidyl transferase cavity:

two oxygens of the para-nitro (p-NO₂) group, the 1OH group, the 3OH group and the 4' carboxyl group.

One of the oxygens of the p-NO₂ group of chloramphenicol appears to form hydrogen bonds with N4 of C2431Dr (C2452Ec) (see Methods for definition), which has been shown to be involved in chloramphenicol resistance¹⁴. The other oxygen of the p-NO₂ group interacts with O2' of U2483Dr (U2504Ec) (Fig. 1a–c).

The 1OH group of chloramphenicol is located at hydrogen-bonding distance (less than 4 Å) to N2 of G2044Dr (G2061Ec) of the 23S rRNA. This nucleotide has been implicated in chloramphenicol resistance in rat mitochondria¹⁴. A mutation of the neighbouring nucleotide, A2062Ec, confers resistance to chloramphenicol in *Halobacterium halobium*¹⁸.

The 3OH group of chloramphenicol is fundamental for its activity, and *in vivo* modification of this group confers resistance to the drug^{19,20}. The 3OH group is within hydrogen-bonding distance to 4'O of U2485Dr (U2506Ec) and to O2' of G2484 (G2505Ec). The 3OH group of chloramphenicol is also involved in interactions coordinated through a putative Mg ion (Mg-C1, see

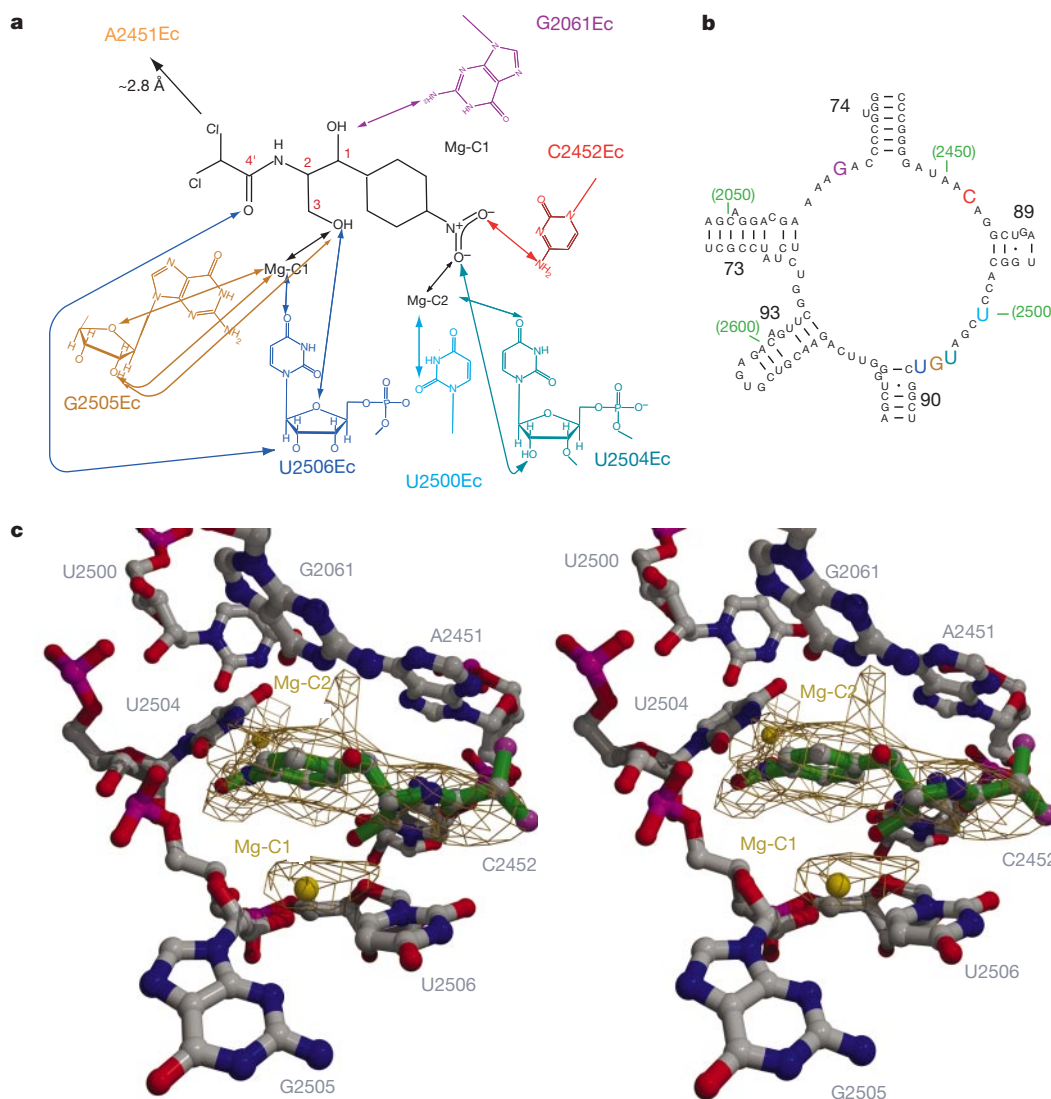


Figure 1 Interaction of chloramphenicol with the peptidyl transferase cavity. **a**, Chemical structure diagram of chloramphenicol showing the interactions (arrows) of its reactive groups with the nucleotides of the peptidyl transferase cavity. Arrows between two chemical moieties indicate that the two groups are less than 4.4 Å apart. **b**, Secondary structure of the peptidyl transferase ring of *D. radiodurans* showing the nucleotides involved in the interaction with chloramphenicol (coloured nucleotides). The colours in the

secondary structure diagram match those of the chemical diagram. **c**, Stereo view showing the nucleotides interacting with chloramphenicol at the peptidyl transferase cavity of *D. radiodurans*. The difference electron density map ($2F_o - F_c$) is contoured at 1.2σ . The antibiotic is shown in green. Nucleotide numbering is according to the *E. coli* sequence. Putative Mg ions (Mg-C1, Mg-C2) are indicated.

below). The 4' carbonyl group of chloramphenicol appears to form a hydrogen bond with the 2'OH of U2485Dr (U2506Ec) (Fig. 1a).

Mg²⁺ antibiotic interactions

In addition to the interactions between chloramphenicol and RNA nucleotides, two ions (Mg-C1 and Mg-C2) are involved in chloramphenicol binding. We propose these ions to be hydrated Mg²⁺. However, at the present resolution we cannot exclude the possibility of the ions being either K⁺ or NH₄⁺. These ions are not present in the native 50S structure or in the 50S complexes with the other antibiotics that we analysed. Thus, their presence at these particular locations depends on chloramphenicol binding.

Mg-C1 mediates the interaction of the 3OH group of chloramphenicol with O4 of U2485Dr (U2506Ec) and with the 2'OH and 4'O of G2484Dr (G2505Ec). Both of these nucleotides have previously been reported as being protected by chloramphenicol¹⁶. Mg-C2 mediates the interaction of one of the oxygens of the p-NO₂ group with O2 of U2479Dr (U2500Ec) and O4 of U2483Dr

(U2504Ec). Mg-C1- and Mg-C2-mediated interactions further stabilize the interaction of chloramphenicol with the peptidyl transferase cavity. The presence of these ions seems crucial for the interaction of chloramphenicol with the ribosome, and thus for the inhibition mechanism of this drug.

Notably, the putative Mg²⁺ (Mg101) found overlapping with the chloramphenicol location in the native structure is not observed in the chloramphenicol–50S complex, suggesting that the coordinating effect of chloramphenicol is sufficient to maintain the local structure of the 50S subunit in the absence of Mg101. The involvement of metal ions in chloramphenicol binding can partially explain why chloramphenicol, in spite of being a small molecule, has been shown by chemical footprinting to protect certain nucleotides on the peptidyl transferase ring from chemical attack, or to enhance the chemical reactivity of other nucleotides in the same region¹⁴. Furthermore, generation of new metal-ion-binding sites owing to antibiotic binding, as we observed for chloramphenicol, may explain the mode of action and/or binding of other drugs as well, and could be used as a tool in rational drug design.

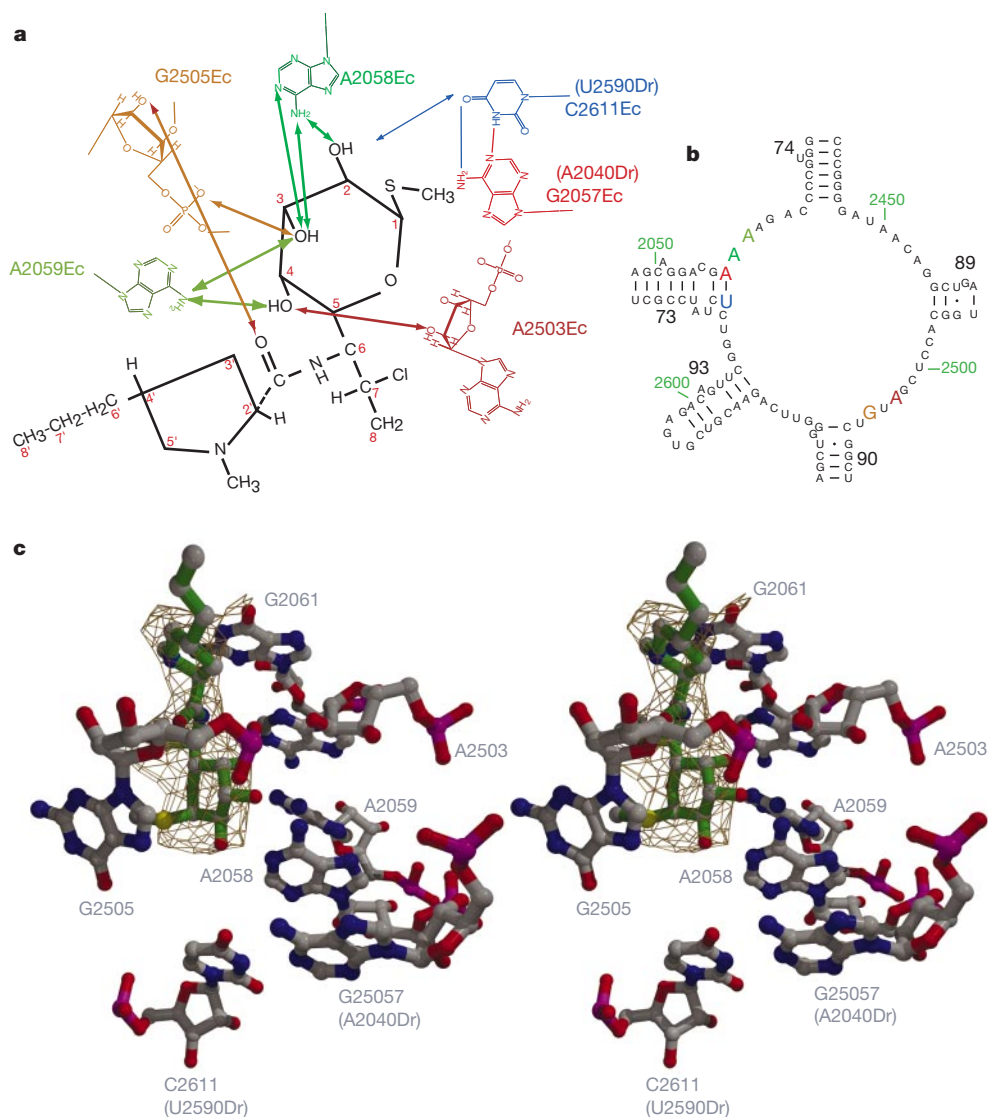


Figure 2 Interaction of clindamycin with the peptidyl transferase cavity. **a**, Chemical structure diagram of clindamycin showing the interactions (arrows) of its reactive groups with the nucleotides of the peptidyl transferase cavity. Arrows between two chemical moieties indicate that the two groups are less than 4.4 Å apart. **b**, Secondary structure of the peptidyl transferase ring of *D. radiodurans* showing the nucleotides involved in the

interaction with clindamycin (coloured nucleotides). The colours in the secondary structure diagram match those of the chemical diagram. **c**, Stereo view showing the nucleotides interacting with clindamycin at the peptidyl transferase cavity of *D. radiodurans*. The difference electron density map ($2F_o - F_c$) is contoured at 1.2σ . The antibiotic is shown in green. Nucleotide numbering is according to the *E. coli* sequence.

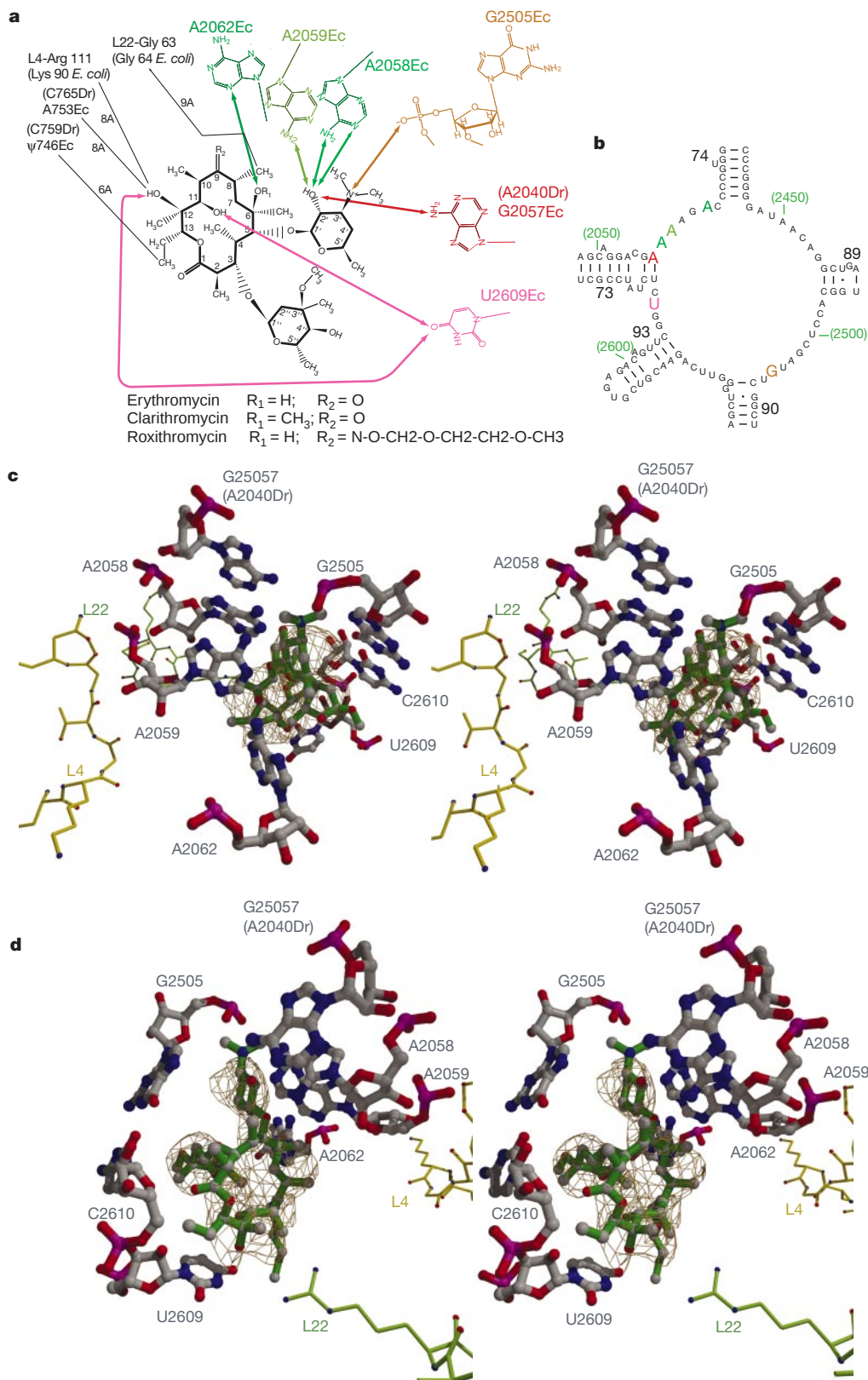


Figure 3 Interaction of macrolides with the peptidyl transferase cavity. **a**, Chemical structure diagram of the macrolides (erythromycin, clarithromycin and roxithromycin) showing the interactions (coloured arrows) of the reactive groups of the macrolides with the nucleotides of the peptidyl transferase cavity (coloured). Coloured arrows between two chemical moieties indicate that the two groups are less than 4.4 Å apart. Distance of macrolide moieties to groups implicated previously in antibiotic interaction (L4, L22 and domain II of the 23S rRNA) are shown. **b**, Secondary structure of the peptidyl transferase ring of *D. radiodurans* showing the nucleotides involved in the interaction with clindamycin

(coloured nucleotides). The colours in the secondary structure diagram match those of the chemical diagram. **c**, Stereo view showing the erythromycin binding site at the peptidyl transferase cavity of *D. radiodurans*. The stereo view for clarithromycin is identical to that of erythromycin. **d**, Stereo view showing the roxithromycin-binding site at the peptidyl transferase cavity of *D. radiodurans*. For **c** and **d** the difference electron density map ($2F_o - F_c$) is contoured at 1.2σ . Green, antibiotic; yellow, part of ribosomal protein L4; light green, part of ribosomal protein L22. Nucleotides that interact with the antibiotic are shown with their chemical structure. Nucleotide numbering is according to the *E. coli* sequence.

Clindamycin

In contrast to chloramphenicol, lincosamides interact with the A and P site²¹. Although the binding site for the lincosamide clindamycin differs from that of chloramphenicol at the peptidyl transferase cavity the two sites appear to be partially overlapping (Figs 2a–c and 4). We did not localize any new metal ions involved in the binding of clindamycin; however, Mg101 was also displaced on clindamycin binding.

Clindamycin has three hydroxyl groups in its sugar moiety that can participate in hydrogen-bond formation (Fig. 2a–c). The 2OH of clindamycin appears to form a hydrogen bond with N6 of A2041Dr (A2058Ec), the pivotal nucleotide for lincosamide binding²². Although the 2OH group is also less than 4.0 Å away from O6 of A2590Dr (G2611Ec), additional hydrogen bonds to this nucleotide is unlikely, because mutations of U2590Dr (C2611Ec) appear to alter only the conformation of the 23S rRNA and thus affect the position of nucleotide A2041Dr (G2058Ec)²³ (Fig. 2c) (see below also).

The 3OH group can interact with N6 of nucleotide A2041Dr (A2058Ec) and with the non-bridging phosphate-oxygens of G2484Dr (G2505Ec). Thus, N6 of A2041Dr (A2058Ec) can interact with both the 2OH group and the 3OH group of clindamycin. These structural data explain in the most straightforward way why A2041Dr (A2058Ec) mutations confer resistance to clindamycin. The hydrogen bond with N6 of A2041Dr (A2058Ec) can also explain why the dimethylation of the N6 group, which disrupts the hydrogen bonds, causes resistance to lincosamides²⁴. The 3OH group of clindamycin could additionally interact with N1 of A2041Dr (A2058Ec) and N6 of A2042 (A2059Ec).

The 4OH group of clindamycin could form a hydrogen bond with N6 of A2042 (A2059Ec). This interaction explains why mutations in A2042 (A2059Ec) cause clindamycin resistance in several bacterial pathogens²⁴. In addition to the sugar-mediated interactions, the carbonyl group of clindamycin could form a hydrogen bond with 2'OH of G2484 (G2505Ec).

The A- and P-site character of clindamycin may be explained by the location of its proline moiety and the interaction of its 8' carbon and its sulphur atom with the 23S rRNA. The location of the proline moiety of bound clindamycin partially overlaps that of the phenyl moiety of chloramphenicol. The proline moiety is on average 3.5 Å away from the tyrosyl moiety of the puromycin-binding site described by ref. 2, thus giving an explanation to the A-site character of clindamycin. The 8' carbon of clindamycin is located 2.5 Å away

from the N3 of C2431Dr (C2452Ec), a nucleotide shown to be involved in P-site tRNA binding²⁵. This location of the 8' carbon may explain the P-site character of clindamycin. The sulphur atom of clindamycin is located around 3 Å from base G2484Dr (G2505Ec), which has been photo-crosslinked to the tRNAs bound to the A and P sites²⁵. Although the two latter clindamycin groups cannot form hydrogen bonds, interactions, such as van der Waals or hydrophobic interactions, between these groups and nucleotides of the 23S rRNA may be expected.

Macrolides

In contrast to chloramphenicol and the lincosamides, macrolides of the erythromycin class do not block peptidyl transferase activity²⁶. Although they bind to the peptidyl transferase ring, the erythromycin group of the macrolides, which includes clarithromycin and roxithromycin, is thought to block the tunnel that channels the nascent peptides away from the peptidyl transferase centre^{2,27,28}. Our results confirm this assumption. We could unambiguously determine that the macrolides erythromycin, clarithromycin and roxithromycin all bind to the same site in the 50S subunit of *D. radiodurans*, at the entrance of the tunnel (Fig. 5). Their binding contacts clearly differ from those of chloramphenicol, but overlap those of clindamycin to a large extent (Figs 3a and 4).

Most of the 14-member-ring macrolides, which include erythromycin and its related compounds, have three structural components: the lactone ring, the desosamine sugar, and the cladinose sugar. The reactive groups of the desosamine sugar and the lactone ring mediate all the hydrogen-bond interactions of erythromycin and its second generation derivatives clarithromycin²⁹ and roxithromycin³⁰ with the peptidyl transferase cavity.

The 2'OH group of the desosamine sugar appears to form hydrogen bonds at three positions: N6 and N1 of A2041Dr (A2058Ec) and N6 of A2042Dr (A2059Ec). The hydrogen bonds between the 2'OH and N1 and N6 of A2041Dr (A2058Ec) explain why this nucleotide is essential for macrolide binding, thus shedding light on the two most common ribosomal resistance mechanisms against macrolides: the N6 dimethylation of A2041Dr (A2058Ec) by the erythromycin resistance family of methylases³¹, and rRNA mutations changing nucleotide identity at this position³². The dimethylation of the N6 group would not only add a bulky substituent causing steric hindrance for the binding, but will prevent the formation of hydrogen bonds to the 2'OH group. The

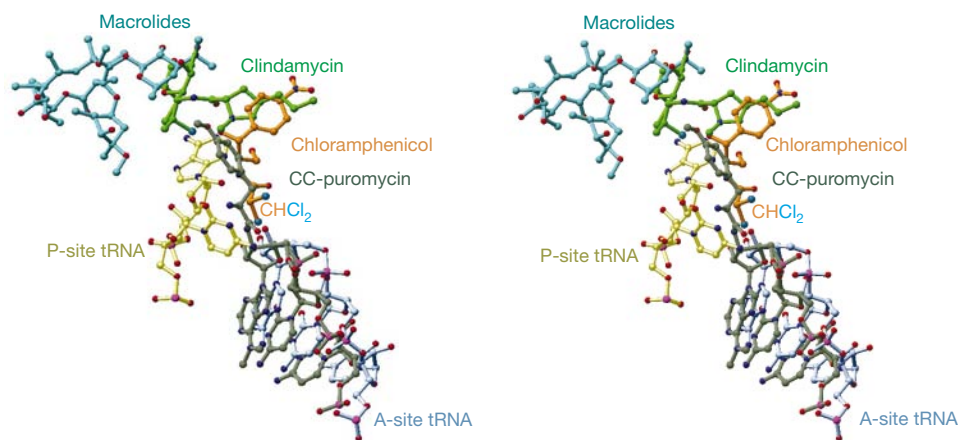


Figure 4 Relative position of chloramphenicol, clindamycin and macrolides with respect to CC-puromycin and the 3'-cytosine-adenine (CA) end of P-site and A-site tRNAs.

The location of CC-puromycin was obtained by docking the position reported by ref. 2 into the peptidyl transferase centre of *D. radiodurans*. The location of the 3'-CA end of P- and A-site tRNAs was obtained by docking the position reported by ref. 42 into the

peptidyl transferase centre of *D. radiodurans*. Light blue, 3'-CA end of A-site tRNA; light yellow, 3'-CA end of P-site tRNA; grey, puromycin; gold, chloramphenicol; green, clindamycin; cyan, macrolides (erythromycin). Oxygen atoms are shown in red and nitrogen atoms in dark blue. CHCl₂ indicates the location of the dichloromethyl moiety of chloramphenicol.

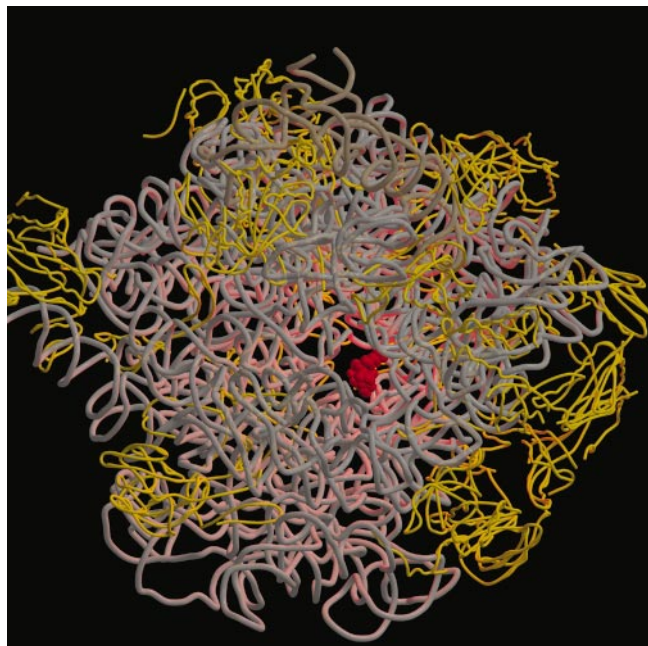


Figure 5 Top view of the *D. radiodurans* 50S subunit showing erythromycin (red) bound at the entrance of the tunnel. Yellow, ribosomal proteins; grey 23S rRNA; dark grey, 5S rRNA.

mode of interactions proposed by our structure implies that a mutation at A2041Dr (A2058Ec) will disrupt the pattern of hydrogen bonding, therefore impairing binding and rendering bacteria resistant. A2041Dr (A2058Ec) is one of the few nucleotides of the peptidyl transferase ring that are not conserved among all phylogenetic domains. Mitochondrial and cytoplasmic rRNAs of higher eukaryotes have a guanosine at position 2058Ec³³. Therefore, the proposed mode of interaction explains the selectivity of macrolides for bacterial ribosomes.

Our results also explain the importance of the 2'OH group of the desosamine sugar for erythromycin binding known from structure activity relation (SAR) studies³⁴. Changing the 2'OH group will not allow the suggested mode of interaction to take place. The 2'OH of the desosamine sugar is located 4 Å away from N6 of A2040Dr (G2057Ec), a nucleotide forming the last base pair of helix 73. Although an interaction with this group cannot be ruled out, this base pair seems to be engaged mainly in maintaining the right conformation of A2041Dr (A2058Ec)²³. Moreover, in contrast to most eubacteria where the last base pair of helix 73 is a G-C pair, in *D. radiodurans* this base pair is an A-U pair, thus making a specific interaction of A2040Dr (G2057Ec) in the binding of macrolides unlikely.

The dimethylamino group of the desosamine sugar exists in both protonated (>96–98%) and neutral (2–4%) forms. This group, if protonated, could interact through ionic interactions in a pH-dependent manner with the backbone oxygen of G2484Dr (G2505Ec) (Fig. 3a, b). At physiological pH, both species appear to be able to bind ribosomes with the same kinetics³⁴; however, the only study that revealed a strong correlation between pH and potency of inhibition of *in vitro* protein synthesis for the macrolide erythromycin concludes that the neutral macrolide form was the inhibitory species³⁵. If the neutral macrolide form is in fact the inhibitory species, a hydrogen bond between the dimethylamine of the macrolides and the nucleotide G2484Dr (G2505Ec) could not be formed.

On the basis of our structural information, it will be of interest to derivatize macrolides at the dimethylamine position. These modifications could improve binding and result in the inhibition of the

peptidyl transferase activity. One of the few macrolides able to inhibit peptidyl transferase is tylosin. Tylosin, which contains a mycarose moiety, has been shown to protect A2506Ec, the neighbouring nucleotide of G2484Dr (G2505Ec), from chemical modification³⁶. Similarly, modification of the dimethylamine group to a reactive group that would interact with G2484Dr (G2505Ec) in a pH-independent manner may lead to more effective macrolides and structurally related compounds.

Only three of the hydroxyl moieties of the lactone ring are within hydrogen-bonding distance to the 23S rRNA: the 6-OH, the 11-OH and the 12-OH groups. The 6-OH group is within hydrogen-bonding distance to N6 of A2045Dr (A2062Ec). Although the 6-OH group has been substituted by a methoxy group to improve acid stability in clarithromycin²⁹, the oxygen of the methoxy group is still at hydrogen-bonding distance to N6 of A2045Dr (A2062Ec). Our structure explains why the chemical footprinting experiments implicate this nucleotide in macrolide binding³⁷.

The 11-OH and 12-OH groups of the lactone ring could each form a hydrogen bond with O4 of U2588Dr (U2609Ec). Hydrogen bonds at these positions can explain substitutions of any of these two hydroxyl groups causing a moderate decrease of binding, as would be expected for groups involved in only one hydrogen-bond interaction with the same 23S rRNA nucleotide³⁵.

The reactive groups of the cladinose sugar are not involved in hydrogen-bond interactions with the 23S rRNA. Cladinose dispensability is confirmed by SAR studies³⁵ showing that the 4'-OH is dispensable for macrolide binding. A closely related group of antibiotics, the ketolides, which bind tighter to ribosomes than macrolides, do not have a cladinose sugar. Moreover, the K_d of erythromycin and the ketolide RU56006, a derivative of erythromycin that lacks the cladinose sugar, are on the same order of magnitude³⁷.

Although the three macrolides differ slightly in structure (Fig. 3a), we observed no differences in their binding site or mode. Our electron density map shows a small part of the etheroxime chain (R2 in Fig. 3a) of roxithromycin pointing towards the tunnel (Fig. 3d); however, this part is not involved in interactions with 23S rRNA or ribosomal proteins.

Our structure does not indicate direct interactions between erythromycin and helix 35 (H35), which has previously been implicated in the binding of erythromycin^{37,38}, and resistance mutations³⁷. The distance between OH-11 of the lactone ring and N4 of C765Dr (A752Ec), a nucleotide of H35 proposed to interact with erythromycin^{37,38}, is 8.5 Å. The CH₃ group branching from position 13 of the lactone ring is located about 4.5 Å away from the O2 of C759Dr (G745Ec) of H35. Nonetheless, hydrophobic van der Waals or ion-coordinated interactions between the lactone ring and this helix loop of the 23S rRNA cannot be ruled out, suggesting that footprinting effects at A752Ec might be of an allosteric nature.

Two ribosomal proteins, L4 and L22, have been implicated in erythromycin resistance³⁹. The closest distance of erythromycin (12-OH) to L4 (R111Dr/K90Ec) is 8 Å, whereas the closest distance from (8-CH₃) to L22 (G63Dr/G64Ec) is 9 Å (Fig. 3a). These distances are greater than one would expect for any meaningful chemical interaction. Therefore, the resistance of macrolides acquired by mutations in these two proteins is probably the product of an indirect effect that is produced by a perturbation of the 23S rRNA owing to the mutated proteins⁴⁰.

Overall, it is difficult to explain the high affinity of macrolides (K_d 10⁻⁸ M) for the ribosome solely by the proposed seven hydrogen bonds. The binding of erythromycin and its derivatives may be further stabilized by van der Waals forces, hydrophobic interactions and the geometry of the rRNA that tightly surrounds the macrolide molecules.

Overlapping binding sites

Two of the 23S rRNA nucleotides involved in the interactions with

chloramphenicol are also involved in interactions with clindamycin: G2044Dr (G2061Ec) and G2484Dr (G2505Ec). Both nucleotides were previously shown by chemical footprinting to be protected by chloramphenicol or clindamycin binding¹⁵.

The common nucleotide moieties involved in the interaction of the 23S rRNA with clindamycin and with the macrolides are N6 of A2041Dr (A2058Ec), N6 of A2042Dr (A2059Ec) and the non-bridging phosphate oxygen of U2484Dr (U2505Ec). These positions are targeted by the sugar moiety of clindamycin and the desosamine sugar of the macrolides, both of which are located in almost identical locations in the structure of the 50S–clindamycin and the 50S–macrolide complexes. The overlapping of binding sites may explain why clindamycin and macrolides bind competitively to the ribosome and why most RNA mutations conferring resistance to macrolides also confer resistance to lincosamides¹⁵.

All examined antibiotics target nucleotide G2484Dr (G2505Ec). The importance of this nucleotide position has been established previously⁴¹. Although this nucleotide was shown to be protected from chemical modification on chloramphenicol, lincosamide, or macrolide binding^{16,17,22}, no mutation of G2484Dr (G2505Ec) conferring resistance to these antibiotics has been reported. In addition, G2484Dr (G2505Ec) is also one of the nucleotides protected on binding of peptidyl-tRNA⁴². The identity of this nucleotide is important for protein synthesis, albeit not for ribosome–antibiotic interactions where the position of its backbone oxygen or the 2'OH of the sugar appear to be the essential requirement.

Although overlapping binding sites may be coincidental, common nucleotide moieties may also point towards essential 23S rRNA nucleotides targeted by antibiotics to inhibit ribosomal function. The structural information derived from the overlapping binding sites will not only be useful to understand antibiotic action but could also have therapeutic implications as it can indicate which antibiotic combination therapies would be meaningful, and can open the door for the design of hybrid antibiotics.

Mechanism of antibiotic action

The relative binding sites for chloramphenicol, clindamycin and erythromycin at the 23S rRNA with respect to the A-site substrate analogue CC-puromycin (according to ref. 2) and the 3' end of P- and A-site tRNAs (according to ref. 43) docked in our structure are shown in Fig. 4. From these sites it can be seen that chloramphenicol acts as a competitor of puromycin and thus as an inhibitor of the A site, consistent with previous findings²⁶. In contrast to puromycin, which acts as a structural analogue of the 3'-end aminoacyl tRNA, the location of chloramphenicol suggests that this drug may act by interfering with the positioning of the aminoacyl moiety in the A site. Thus, chloramphenicol may physically prevent the formation of the transition state during peptide-bond formation. Furthermore, the carbon of the dichloromethyl moiety of chloramphenicol, a moiety shown to be important for chloramphenicol activity⁴⁴, is close to the amino acceptor group of CC-puromycin² (Fig. 4). The presence of such an electronegative moiety in the vicinity of the amino acceptor could also hamper or at least contribute to the inhibitory effect of chloramphenicol on peptide-bond formation.

Clindamycin clearly bridges the binding site of chloramphenicol and that of the macrolides (Fig. 4), and overlaps directly with the A and P sites. Thus, its binding position gives a structural explanation for the hybrid nature of its A and P sites. Clindamycin can interfere with the positioning of the aminoacyl group at the A site and the peptidyl group at the P site while also sterically blocking the progression of the nascent peptide towards the tunnel.

The macrolide binding site is at the entrance of the tunnel (Fig. 5). This location can allow the formation of 6–8 peptide bonds before the nascent protein chain reaches the bound macrolide. When bound, macrolides reduce the diameter of the tunnel entrance from 18–19 Å to less than 10 Å; however, as the space not occupied

by erythromycin hosts a putative, hydrated Mg²⁺ ion (Mg33), the passage available for the nascent protein will be between 6 and 7 Å in diameter. Our structural results are consistent with previous biochemical findings, showing that small peptides (up to eight residues) can be produced by erythromycin-bound ribosomes⁴⁵. It is possible that ribosomes could escape the inhibitory effects of macrolides if proteins with specific leader sequences are translated that allow the threading of the narrowed tunnel.

Overall, the binding sites of the antibiotics analysed suggest that their inhibitory action is not only determined by their interaction with specific nucleotides. These antibiotics could also inhibit peptidyl transferase by interfering with the proper positioning and movement of the tRNAs at the peptidyl transferase cavity. This sterical hindrance may be direct, as in the case of chloramphenicol, or indirect, as in the case of the three macrolides. In addition to sterical hindrance, antibiotic binding may physically link regions known to be essential for the proper positioning of the aminoacyl- and peptidyl-tRNAs, and thus prevent the conformational flexibility needed for protein biosynthesis.

The overall interaction of each analysed antibiotic together with the lack of any major conformational changes on antibiotic binding to the ribosome points towards the theory that the peptidyl transferase centre evolved as a template for the proper binding of the activated substrates, that is, the aminoacyl-tRNA and the peptidyl-tRNA⁴. However, we cannot completely exclude the presence of a catalytic nucleotide.

Discussion

Our work reveals the functional interactions involved in the binding of antibiotics to the peptidyl transferase cavity of a bacterial ribosome. All 23S rRNA nucleotides involved in interactions with these five antibiotics are part of the peptidyl transferase cavity of the 50S subunit. None of the antibiotics examined show any direct interaction with ribosomal proteins. Chloramphenicol targets mainly the A site, where it interferes directly with substrate binding. Clindamycin interferes with the A site and P site substrate binding and physically hinders the path of the growing peptide chain. Macrolides bind at the entrance to the tunnel where they sterically block the progression of the nascent peptide. Although the binding sites of the antibiotics differ from each other, they show some overlapping nucleotides.

The structural model of the peptidyl transferase centre in complex with the examined antibiotics should not only enable a rational approach for antibiotic development and therapy strategies but could be also used to identify new target sites on the eubacterial ribosome. □

Methods

Base numbering

Bases are numbered according to the *D. radiodurans* 23S rRNA sequence. A 'Dr' is added at the end of the base number for further identification. Corresponding *Escherichia coli* base numbers of the 23S rRNA sequence are given in parenthesis and have an 'Ec' added after the base number.

D. radiodurans cell growth and ribosome isolation

Cells were grown as recommended by the American Type Culture Collection (ATCC) with minor modifications to ATCC medium 679. Ribosomes were isolated as described⁴⁶.

50S crystals

Crystals of 50S belong to the space group *I*222 and contain one particle per asymmetric unit. Crystals were grown by vapour diffusion, in 10 mM MgCl₂, 60 mM NH₄Cl, 5 mM KCl, 10 mM HEPES, pH 7.8, using low concentrations (0.1–1%) of poly- and mono-valent alcohols (dimethyl-hexandiol:ethanol) as precipitant. Co-crystallization was carried out in the presence of 0.8–3.5 mM of the antibiotics chloramphenicol, clindamycin, erythromycin and roxithromycin. Soaking of 50S crystals was performed for clarithromycin and chloramphenicol in solutions containing 0.1 mM of the respective antibiotic. Chloramphenicol data was collected from crystals grown in the presence of the antibiotic as well as from soaked crystals. Both data sets resulted in the same experimental density in the chloramphenicol difference maps and thus were merged. Difference maps

for all other tested antibiotics were obtained from either co-crystallized or soaked crystals alone.

X-ray diffraction

Data were collected at 85 K from shock-frozen crystals with bright synchrotron radiation beam at ID19 at Argonne Photon Source/Argonne National Laboratory (APS/ANL), ID14/2 and 4 at European Synchrotron Radiation Facility/European Molecular Biology Laboratory (ESRF/EMBL), and at BW6 at Deutsches Elektronen-Synchrotron (DESY). Data were recorded on MAR345, Quantum 4, or APS-CCD (charge-coupled device) detectors and processed with HKL2000 (ref. 47).

Placements and refinement

The 3.1 Å structure of the 50S subunit was refined against the structure factor amplitudes of each of the 50S–antibiotic complexes, using rigid body refinement as implemented in CNS¹⁸. We used SigmaA weighted difference maps for the initial manual placement of the antibiotics. Each of the 50S–antibiotic models was further refined in Refmac⁴⁹. For the calculation of free *R*-factor (as reported in Table 1) a subset of reflections (10% of the data) was omitted from the refinement.

The experimental density allowed us the unequivocal placement of the five antibiotics that we studied. Taking into account the estimated coordinate error (Table 1) and the limits imposed by the resolution, we analysed all of the rRNA groups located less than 4.4 Å from any antibiotic moiety. Here we report all moieties from the previous group having the potential to form hydrogen bonds.

Coordinates and figures

Coordinates of chloramphenicol, clindamycin, erythromycin, roxithromycin and clarithromycin were taken from Cambridge Structural Database. Antibiotics were modelled into the difference density based on their crystal structure. Figures were produced with RIBBONS⁵⁰ or MOLSCRIPT⁵¹ programs.

Received 3 August; accepted 27 September 2001.

- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
- Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).
- Pioletti, M. *et al.* Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J.* **20**, 1829–1839 (2001).
- Winberry, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).
- Schluenzen, F. *et al.* Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**, 615–623 (2000).
- Polacek, N., Gaynor, M., Yassin, A. & Mankin, A. S. Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature* **411**, 498–501 (2001).
- Thompson, J. *et al.* Analysis of mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proc. Natl Acad. Sci. USA* **98**, 9002–9007 (2001).
- Barta, A. *et al.* Mechanism of ribosomal peptide bond formation. *Science* **291**, 203a (2001).
- Spahn, C. M. T. & Prescott, C. D. Throwing a spanner in the works: antibiotics and the translation apparatus. *J. Mol. Med.* **74**, 423–439 (1996).
- Cundliffe, E. in *The Ribosome: Structure, Function and Evolution* (eds Hill, W. E. *et al.*) 479–490 (ASM, Washington DC, 1990).
- Vester, B. & Douthwaite, S. Macrolide resistance conferred by base substitutions in 23S rRNA. *Antimicrob. Agents Chemother.* **45**, 1–12 (2001).
- Harms, J. *et al.* High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* (in the press).
- Vester, B. & Garrett, R. A. The importance of highly conserved nucleotides in the binding region of chloramphenicol at the peptidyl transfer center of *Escherichia coli* 23S ribosomal RNA. *EMBO J.* **7**, 3577–3588 (1988).
- Polacek, N. in *RNA-binding Antibiotics* (eds Schroeder, R. & Wallis, M. G.) 1–10 (Eurekah.Com, Georgetown, 2000).
- Rodriguez-Fonseca, C., Amis, R. & Garrett, R. A. Fine structure of the peptidyl transferase centre on 23S-like rRNAs deduced from chemical probing of antibiotic-ribosome complexes. *J. Mol. Biol.* **247**, 224–235 (1995).
- Moazed, D. & Noller, H. F. Chloramphenicol, erythromycin, carbomycin and vernamycin B protect overlapping sites in the peptidyl transferase region of 23S ribosomal RNA. *Biochemie* **69**, 879–884 (1987).
- Mankin, A. S. & Garrett, R. A. Chloramphenicol resistance mutations in the single 23S rRNA gene of the archaeon *Halobacterium halobium*. *J. Bacteriol.* **173**, 3559–3563 (1991).
- Izard, T. & Ellis, J. The crystal structures of chloramphenicol phosphotransferase reveal a novel inactivation mechanism. *EMBO J.* **19**, 2690–2700 (2000).
- Shaw, W. V. & Leslie, A. G. W. Chloramphenicol acetyl-transferase. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 363–386 (1991).
- Kalliarafopoulos, S., Kalpaxis, D. L. & Coutsoygeorgopoulos, C. New aspects of the kinetics of inhibition by lincomycin of peptide-bond formation. *Mol. Pharmacol.* **46**, 1009–1014 (1994).
- Douthwaite, S. Interaction of the antibiotics clindamycin and lincomycin with *Escherichia coli* 23S ribosomal-RNA. *Nucleic Acids Res.* **20**, 4717–4720 (1992).

- Vester, B., Hansen, L. H. & Douthwaite, S. The conformation of 23S ribosomal-RNA nucleotide A2058 determines its recognition by the Erme methyltransferase. *RNA* **1**, 501–509 (1995).
- Ross, J. I. *et al.* Clinical resistance to erythromycin and clindamycin in cutaneous propionibacteria isolated from acne patients is associated with mutations in 23S rRNA. *Antimicrob. Agents Chemother.* **41**, 1162–1165 (1997).
- Steiner, G., Kuechler, E. & Barta, A. Photo-affinity labeling at the peptidyl transferase center reveals 2 different positions for the A-site and P-sites in domain-V of 23S ribosomal-RNA. *EMBO J.* **7**, 3949–3955 (1988).
- Vazquez, D. *Inhibitors of Protein Synthesis* (Springer, Berlin, 1975).
- Yonath, A., Leonard, K. R. & Wittmann, H. G. A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction. *Science* **236**, 813–816 (1987).
- Milligan, R. A. & Unwin, P. N. Location of exit channel for nascent protein in 80S ribosome. *Nature* **319**, 693–695 (1986).
- Steinmetz, W. E., Bersch, R., Towson, J. & Pesiri, D. The conformation of 6-methoxyerythromycin-A in water determined by proton NMR-spectroscopy. *J. Med. Chem.* **35**, 4842–4845 (1992).
- Gasc, J. C., Dambrieres, S. G., Lutz, A. & Chantot, J. F. New etheroxime derivatives of erythromycin A: a structure–activity relationship study. *J. Antibiot.* **44**, 313–330 (1991).
- Weisblum, B. Erythromycin resistance by ribosome modification. *Antimicrob. Agents Chemother.* **39**, 577–585 (1995).
- Sigmund, C. D., Ettayebi, M. & Morgan, E. A. Antibiotic-resistance mutations in 16S and 23S ribosomal-RNA genes of *Escherichia coli*. *Nucleic Acids Res.* **12**, 4653–4663 (1984).
- Bottger, E. C., Springer, B., Prammananan, T., Kidan, Y. & Sander, P. Structural basis for selectivity and toxicity of ribosomal antibiotics. *EMBO Rep.* **2**, 318–323 (2001).
- Goldman, R. C., Pesik, S. W. & Doran, C. C. Role of protonated and neutral forms of macrolides in binding to ribosomes from gram-positive and gram-negative bacteria. *Antimicrob. Agents Chemother.* **34**, 426–431 (1990).
- Mao, J. C.-H. & Puttermann, M. The intermolecular complex of erythromycin and ribosome. *J. Mol. Biol.* **44**, 347–361 (1969).
- Poulsen, S. M., Kofoed, C. & Vester, B. Inhibition of the ribosomal peptidyl transferase reaction by the mycrosome moiety of the antibiotics carbomycin, spiramycin and tylosin. *J. Mol. Biol.* **304**, 471–481 (2000).
- Hansen, L. H., Mauvais, P. & Douthwaite, S. The macrolide-ketolide antibiotic binding site is formed by structures in domains II and V of 23S ribosomal RNA. *Mol. Microbiol.* **31**, 623–631 (1999).
- Xiong, L. Q., Shah, S., Mauvais, P. & Mankin, A. S. A ketolide resistance mutation in domain II of 23S rRNA reveals the proximity of hairpin 35 to the peptidyl transferase centre. *Mol. Microbiol.* **31**, 633–639 (1999).
- Wittmann, H. G. *et al.* Biochemical and genetic studies of two different types of erythromycin resistant mutants of *Escherichia coli* with altered ribosomal proteins. *Mol. Gen. Genet.* **127**, 175–189 (1973).
- Gregory, S. T. & Dahlberg, A. E. Erythromycin resistance mutation in ribosomal proteins L22 and L4 perturb the high order structure of 23S ribosomal RNA. *J. Mol. Biol.* **289**, 827–834 (1999).
- Saarma, U., Spahn, C. M. T., Nierhaus, K. H. & Remme, J. Mutational analysis of the donor substrate binding site of the ribosomal peptidyltransferase center. *RNA* **4**, 189–194 (1998).
- Moazed, D. & Noller, H. F. Sites of interaction of the CCA-end of peptidyl-transfer RNA with 23S ribosomal-RNA. *Proc. Natl Acad. Sci. USA* **88**, 3725–3728 (1991).
- Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
- Vince, R., Almqvist, R. G., Ritter, C. L. & Daluge, S. Chloramphenicol binding site with analogues of chloramphenicol and puromycin. *Antimicrob. Agents Chemother.* **8**, 439–443 (1975).
- Tenson, T., DeBlasio, A. & Mankin, A. A functional peptide encoded in the *Escherichia coli* 23S rRNA. *Proc. Natl Acad. Sci. USA* **93**, 5641–5646 (1996).
- Noll, M., Hapke, B., Schreier, M. H. & Noll, H. Structural dynamics of bacterial ribosomes. I. Characterization of vacant couples and their relation to complexed ribosomes. *J. Mol. Biol.* **75**, 281–294 (1973).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55**, 247–255 (1999).
- Carson, M. Ribbons. *Methods Enzymol.* **277**, 493–505 (1997).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

Acknowledgements

We thank A. Hofmann for advice; A. Mankin and F. Triana for critical discussions; and T. Auerbach, H. Bartels, W. S. Bennett, H. Burmeister, C. Glotz, M. Glühmann, H. A. S. Hansen, M. Kessler, M. Laschever, S. Meier, J. Muessig, M. Peretz, M. Pioletti, B. Schmidt, A. Sitka, C. Stamer and A. Vieweger for contributing to different stages of these studies. These studies could not have been performed without the cooperation of the staff of the synchrotron radiation facilities at EMBL and Max Planck Gesellschaft at DESY, ID14/2&4 at EMBL/ESRF and ID19/APS/ANL. Support was provided by the Max-Planck Society, the US National Institute of Health, the German Ministry for Science and Technology and the Kimmelman Center for Macromolecular Assembly at the Weizmann Institute. A.Y. holds the Martin S. Kimmel Professorial Chair.

Correspondence and requests for materials should be addressed to F.F. (e-mail: Franceschi@molgen.mpg.de). Coordinates have been deposited in the Protein Data Bank under accession numbers 1JZX, 1JZY, 1JZZ, 1K00 and 1K01.

Energy spectra of quantum rings

A. Fuhrer*, S. Lüscher*, T. Ihn*, T. Heinzel*§, K. Ensslin*, W. Wegscheider† & M. Bichler‡

* Solid State Physics Laboratory, ETH Zürich, 8093 Zürich, Switzerland

† Angewandte und Experimentelle Physik, Universität Regensburg, 93040 Regensburg, Germany

‡ Walter Schottky Institut, TU München, 85748 Garching, Germany

§ Fakultät für Physik, Universität Freiburg i. Br., 79104 Freiburg, Germany

Quantum mechanical experiments in ring geometries have long fascinated physicists. Open rings connected to leads, for example, allow the observation of the Aharonov–Bohm effect¹, one of the best examples of quantum mechanical phase coherence^{2,3}. The phase coherence of electrons travelling through a quantum dot embedded in one arm of an open ring has also been demonstrated⁴. The energy spectra of closed rings⁵ have only recently been studied by optical spectroscopy^{6,7}. The prediction that they allow persistent current⁸ has been explored in various experiments^{9–11}. Here we report magnetotransport experiments on closed rings in the Coulomb blockade regime¹². Our experiments show that a microscopic understanding of energy levels, so far limited to few-electron quantum dots¹³, can be extended to a many-electron system. A semiclassical interpretation of our results indicates that electron motion in the rings is governed by regular rather than chaotic motion, an unexplored regime in many-electron quantum dots. This opens a way to experiments where even more complex structures can be investigated at a quantum mechanical level.

Quantum ring samples have been fabricated on AlGaAs–GaAs heterostructures containing a two-dimensional electron gas (2DEG) with density $5 \times 10^{11} \text{ cm}^{-2}$ and mobility $900,000 \text{ cm}^2 \text{ V s}^{-1}$ at $T = 4.2 \text{ K}$ only 34 nm below the sample surface. The surface of the heterostructure has been locally oxidized by applying a voltage between the conductive tip of an atomic force microscope (AFM) and the 2DEG¹⁴. The electron gas is depleted below the oxidized regions. This technique has been used in other studies to define high quality quantum dots¹⁵. The details of the fabrication process, which is crucial for the high-electronic quality of the quantum ring, are described in ref. 16. Figure 1a shows an AFM image (taken with an unbiased tip directly after the oxidation process) of the oxide lines defining the quantum ring. The width of the quantum point

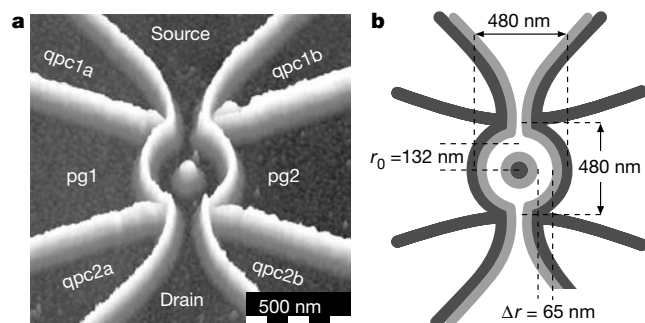


Figure 1 Sample layout. **a**, Micrograph of the quantum ring taken with the unbiased AFM-tip after writing the structure. The oxide lines (bright regions) deplete the 2DEG 34 nm below the surface separating the sample into several conductive (dark) regions. The current is passed from source to drain. The in-plane gates (qpc1a, qpc1b, qpc2a, qpc2b, pg1 and pg2) are used to tune the point contacts and two arms of the ring. **b**, Schematic sketch of the ring. The dark curves represent the oxide lines. From transmission measurements of the point contacts at source and drain we estimate the depletion length to be about 50 nm, which results in an estimated channel width of $\Delta r \approx 65 \text{ nm}$. The average radius of the ring is $r_0 = 132 \text{ nm}$.

contacts connecting the ring to the source (or drain) is controlled by voltages applied to the lateral gate electrodes qpc1a and b (or to qpc2a and b). The number of electrons in the ring can be tuned via the lateral plunger gates pg1 and 2. Shape deformations due to applied in-plane gate voltages are known to be relatively weak^{15,16}. The schematic in Fig. 1b shows the dimensions of the quantum ring.

After the oxidation step the sample has been covered with a metallic top gate electrode. With the combination of in-plane and top gate electrodes the quantum ring can be tuned into the Coulomb blockade regime with the single-particle level spacing being much larger than the thermal energy, kT .

Figure 2b presents a colour plot of the current through the quantum ring as a function of plunger-gate voltage and magnetic field B (applied normal to the 2DEG plane). This measurement was performed at a source–drain voltage $V_{sd} = 20 \mu\text{V}$ and at a temperature of 100 mK in a dilution refrigerator. In Fig. 2a the Coulomb blockade oscillations have been extracted along the horizontal dashed line in Fig. 2b, that is, at constant $B = 92 \text{ mT}$. From corresponding measurements of the Coulomb blockade diamonds we determine a charging energy $E_c = e^2/C_\Sigma \approx 190 \mu\text{eV}$, where C_Σ is the total capacitance of the ring, and e is the elementary charge. The observed discrete level spacings after subtraction of E_c within the constant interaction model¹² can be as large as $\Delta \approx 180 \mu\text{eV}$ (see below). From a simple capacitance model and the ring geometry we estimate that about 200 electrons are distributed on 2 to 3 radial subbands. In Fig. 2b the position as well as the amplitude of the Coulomb blockade peaks oscillate as a function of magnetic field with a period of $\Delta B = 75 \text{ mT}$ (horizontal white lines) which is exactly the Aharonov–Bohm period for this ring. In fact, by opening the point contacts (not shown) we find the well known Aharonov–Bohm oscillations in the conductance with the same ΔB (refs 2, 3, 17 and 18). The oscillations are visualized along a line of constant gate voltage in Fig. 2c (vertical dashed line in Fig. 2b).

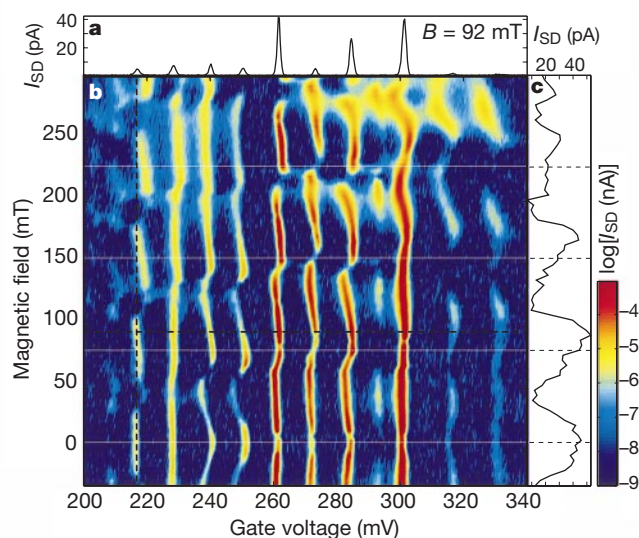


Figure 2 The addition spectrum. **a**, Measurement of Coulomb blockade resonances at fixed magnetic field. The current is measured as a function of a voltage applied to both plunger gates (pg1 and 2) simultaneously. **b**, The evolution of such sweeps with magnetic field results in the addition spectrum shown in colour. The regions of high current (yellow/red) mark configurations in which a bound state in the ring aligns with the Fermi level in source and drain. The Aharonov–Bohm period expected from the ring geometry is indicated by the thin white horizontal lines. **c**, Magnetic field sweep for constant plunger gate voltage $V_{pg} = 218 \text{ mV}$ (dashed line in the colour plot). This peak shows a maximum in amplitude for $B = 0$, whereas other peaks ($V_{pg} = 270 \text{ mV}$) display a minimum.

What determines the magnetic field dependence of positions and amplitudes of the Coulomb blockade resonances? We start the discussion from the energy spectrum of an infinitely thin perfect ring of radius r_0 enclosing m flux quanta⁸:

$$E_{m,l} = \frac{\hbar^2}{2m^*r_0^2}(m+l)^2$$

Here m^* is the mass of the particle and l is the angular momentum quantum number. For a given angular momentum state the energies as a function of magnetic field (or flux m in units of the flux quantum h/e , where h is Planck's constant) lie on a parabola with its apex at $m = -l$, as depicted in Fig. 3a. According to this simple picture a single Coulomb blockade peak should oscillate as a function of B along a zig-zag line (blue curve) with the Aharonov–Bohm period ΔB . Comparison with the measurement in Fig. 2b shows that some peaks indeed move along a zig-zag line, but others show barely any B -dependence, a behaviour which we will discuss later.

We first take a closer look at the h/e -periodic modulation of the Coulomb blockade peak amplitude. A peak follows a line of constant electron number (blue line in Fig. 3a). The current is successively carried by states $(l), (-l-1), (l-1), (-l-2), (l-2)\dots$ when B is increased from zero, that is, a change in state occurs every half flux quantum. The amplitude of Coulomb blockade peaks is determined by the wavefunction overlap between the confined states in the ring and the extended states in source and drain¹². However, the wavefunctions in an infinitely thin perfect ring are independent of magnetic field and given by

$$\Psi_{m,l}(\Phi) = \frac{1}{\sqrt{2\pi}} e^{il\Phi}$$

where Φ is the azimuthal coordinate in the ring. The (lateral) overlap (proportional to $|\Psi|^2$) with source and drain is the same for all states. This model therefore does not predict the observed h/e -periodic modulation of the peak amplitude.

A more realistic but still analytically soluble model takes the finite extent of the wavefunctions in radial direction into account¹⁹, which leads to several radial channels indexed by the quantum number n . Using this model and analysing the Coulomb diamonds in detail, we obtain agreement with the previous estimate of about 2–3 radial modes and about 200 electrons in the ring. For small (large) n , the occupied states at the Fermi level have a larger (or smaller) angular momentum quantum number l at $B = 0$ and consequently display a stronger (or weaker) magnetic field dispersion. This model predicts that the exponential decay of the wave functions in radial direction depends on the value of l but is relatively insensitive to the value of m (for small m). As all states move in zig-zag lines with an h/e periodicity in magnetic field, crossings of states with different angular momenta l may lead to a different wave function overlap and therefore to a modulation of the corresponding Coulomb peak amplitude.

The angular uniformity of the probability density in a perfect ring stems from the cylindrical symmetry which, for the real sample, will be broken by the pure presence of source and drain, by dopants and by the limits of the fabrication procedure. This leads to pinning of the wavefunction and hence to a distinct amplitude of the probability density at source and drain. The perturbation will become especially important at the degeneracy points of levels where it leads to anti-crossing behaviour²⁰. In the simplest case the probability density changes from a uniform to a sinusoidal angle-dependence at the degeneracy points. In this picture, the Aharonov–Bohm-periodic oscillation of the amplitude along a single Coulomb peak can be understood in terms of changing contributions of single particle levels to the current-carrying state.

We now turn to the analysis of the experimental Coulomb blockade peak positions. They are obtained from measurements like the ones shown in Fig. 2b by converting the gate-voltage axis

into an energy scale using the appropriate lever arm as determined from the analysis of the Coulomb blockade diamonds²¹. A constant charging energy of $190 \mu\text{eV}$ is subtracted from the position of a Coulomb maximum^{12,15} and the resulting energies are plotted in Fig. 3c as a function of magnetic field. Clearly many of the peaks move in pairs (see, for example, the black–purple, the blue–red and the green–purple pairs), previously identified as spin pairs¹⁵. The exchange-related spin-splitting energy with a value of around $20 \mu\text{eV}$ for these peaks is on average smaller than the discrete energy level spacing Δ . Electrons therefore successively populate these orbital states with spin-up and spin-down electrons. For other peaks (green, yellow, red) spin pairing is not clearly observed. As depicted in Fig. 3c the orbital states move up and down in the

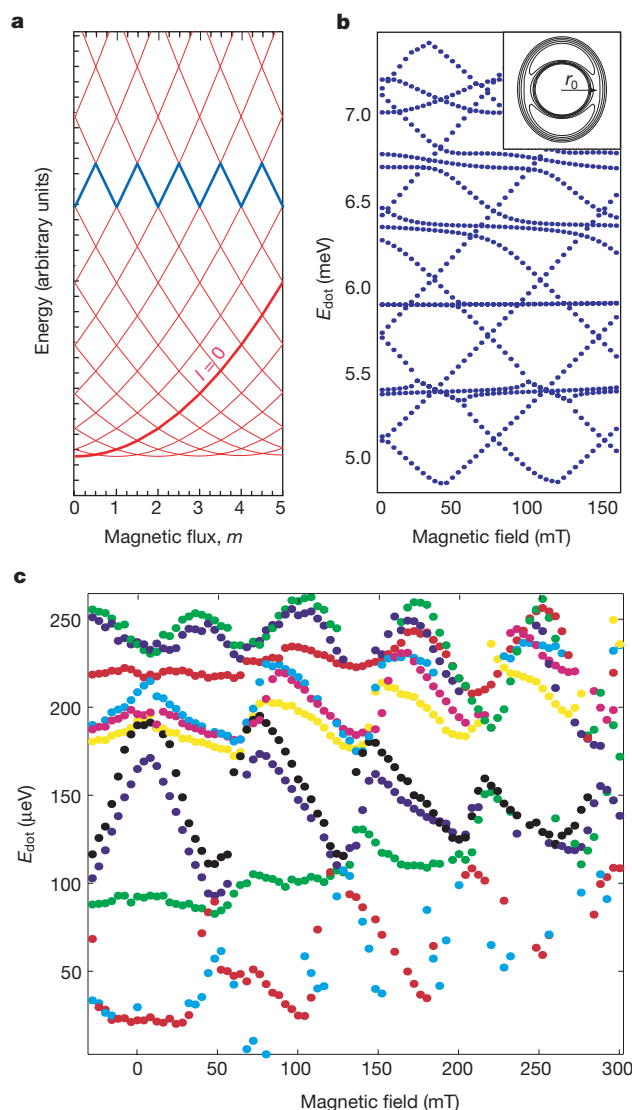


Figure 3 Reconstruction of the energy spectrum. **a**, Theoretical spectrum of a single mode ring. The parabolas with constant l (bold red line corresponding to $l = 0$) have a minimum at $l = m$. The blue zig-zag line corresponds to a Coulomb peak after the charging energy has been subtracted in the constant interaction model. **b**, Calculated spectrum with ring parameters typical for our dot. We assume a slightly asymmetric potential shown in the inset, which mixes states of positive and negative angular momenta. This leads in some cases to eigenenergies that barely depend on magnetic field. **c**, Reconstruction of the energy spectrum of the ring from the data shown in Fig. 2. The plunger gate voltage was converted into dot energy E_{dot} using measurements of the Coulomb diamonds and a constant charging energy of $190 \mu\text{eV}$ was subtracted.

magnetic field with the Aharonov–Bohm period ΔB . In this respect our experiments show the long-predicted energy spectrum characteristic for quantum rings⁵.

Let us look at the states in Fig. 3c that have a very small magnetic field dispersion. In the framework of the model of ref. 19 such ‘flat’ states can occur at the onset of the occupation of the next highest radial channel. However, in our experiment we observe such states over wide ranges of gate voltages and a more refined model is necessary. In Fig. 3b we show a calculation with ring parameters typical for our dot. The spectrum is obtained from the diagonalization of a truncated hamiltonian matrix expressed in the eigenstate basis of ref. 19, but including diagonal and off-diagonal elements given by a symmetry-breaking potential. The angle-independent part ω_0 of ref. 19 is replaced by $\omega_0(1 + \epsilon \cos(2\phi))$ following the idea of shape deformation in ref. 22. This breaks the symmetry, leading to inter- as well as intra-subband coupling. The values used for the calculation were $\hbar\omega_0 = 1.8$ meV, $r_0 = 132$ nm and $\epsilon = 0.06$. Two radial modes are taken into account leading to two sets of parabolic energy dispersions as a function of magnetic field in the unperturbed system. The perturbation mixes states of positive and negative angular momenta—this leads in some cases to eigenenergies that barely depend on magnetic field. Such states can intersect the diamonds formed by strongly oscillating levels in close resemblance to the experimental findings. The model obviously can only give the general tendency of the experimental spectra. Nevertheless, the dominant deviations from the perfect ring spectrum can be understood as the result of a symmetry-breaking potential perturbation.

We estimate the contribution of a particular strongly oscillating state to the persistent current⁸ from the experimental dispersion (that is, black dots in Fig. 3c) and obtain a value of 5 nA. If we assume that the currents of all the lower lying states approximately add up to zero this current is also an estimate of the total persistent current in the ring and the value is consistent with previous magnetisation measurements^{9,10,11}.

For quantum dots containing a small number of electrons, the

shell structure of the level occupancy can clearly be detected because of the dominating cylindrical symmetry¹³. Quantum dots containing many electrons are usually described in the context of an underlying classically chaotic geometry, since small perturbations of parameter space such as potential shape or magnetic field can induce parametric fluctuations in the energy levels and consequently in the Coulomb peak positions. Several theoretical and experimental works investigate whether the spectra of such many-electron quantum dots can adequately be described by random matrix theory²³. Our quantum ring represents a many-electron Coulomb blockaded system with regular geometry. This, together with the small number of radial modes, is why we can qualitatively understand the principal features of the observed energy spectrum without the need of a statistical analysis. The source of this observation lies in the circular geometry of our ring.

To support this view we fabricated a square quantum dot with a circular antidot in the centre. This system is considered a Sinai billiard and is one of the theoretically best-studied examples of a classically chaotic system. Figure 4 shows the evolution of the conductance as a function of plunger gate and magnetic field of this system presented in a way comparable to Fig. 2b for the quantum ring. Around $B = 0$ the Coulomb peak maxima fluctuate irregularly as a function of magnetic field with an average period compatible with an Aharonov–Bohm-type argument. As the magnetic field is increased to a value where the classical cyclotron diameter matches the antidot’s circumference, we observe pronounced B -periodic behaviour of the Coulomb peak maxima in amplitude and position. As in antidot lattices²⁴, the magnetic field is expected to induce regular parts in the predominantly chaotic phase space existing at $B = 0$.

The detailed analysis of quantum rings demonstrates that even in many-electron Coulomb blockaded systems a detailed understanding of the energy spectrum can be obtained. With advanced fabrication techniques we will be able to understand more complex and multiply connected structures on a quantum mechanical level. Electron–electron interactions beyond the constant interaction model are believed to play a minor role in our quantum ring. One indication is the observation of spin pairs and relatively small spin splitting. Once ring structures with only one radial mode occupied are available, such quantum rings could be used to investigate spin effects²⁵ or even Luttinger liquid behaviour in a circular one-dimensional system with periodic boundary conditions. We could also search for persistent current effects in the transport signatures of a quantum ring now that the energy spectrum is experimentally accessible. □

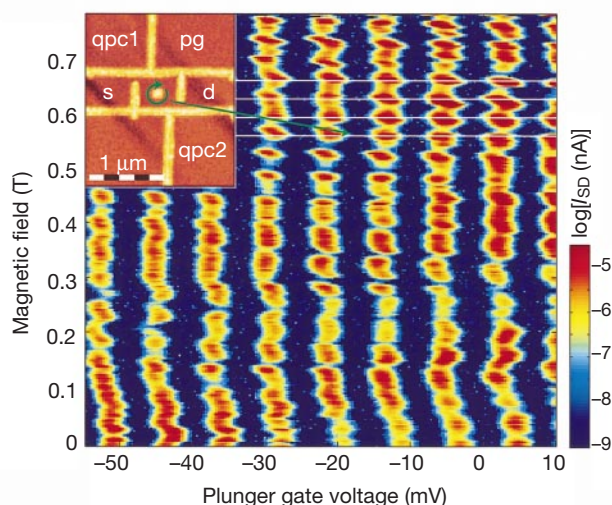


Figure 4 Addition spectrum of a Sinai billiard. The inset shows an AFM micrograph of the square quantum dot with a circular antidot in the centre. The colour plot is similar to the one in Fig. 2b and shows the evolution of the conductance as a function of plunger gate voltage and magnetic field. Around $B = 0$ the Coulomb maxima fluctuate irregularly as a function of magnetic field. As the magnetic field is increased to a value where the classical cyclotron orbit matches the antidot’s circumference (indicated by the green circle and the arrow in the inset), we observe pronounced $\hbar/e$ -periodic behaviour (white lines) in amplitude and position of the Coulomb peaks, indicating quenching of the chaotic behaviour.

Received 13 July; accepted 5 September 2001.

1. Aharonov, Y. & Bohm, D. Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.* **115**, 485–491 (1959).
2. Webb, R. A., Washburn, S., Umbach, C. P. & Laibowitz, R. B. Observation of $\hbar/e$ Aharonov–Bohm oscillations in normal-metal rings. *Phys. Rev. Lett.* **54**, 2696–2699 (1985).
3. Timp, G. *et al.* Observation of the Aharonov–Bohm effect for $\omega_c\tau > 1$. *Phys. Rev. Lett.* **58**, 2814–2817 (1987).
4. Schuster, R. *et al.* Phase measurement in a quantum dot via a double-slit interference experiment. *Nature* **385**, 417–420 (1997).
5. Byers, N. & Yang, C. N. Theoretical considerations concerning quantized magnetic flux in superconducting cylinders. *Phys. Rev. Lett.* **7**, 46–49 (1961).
6. Lorke, A. *et al.* Spectroscopy of nanoscopic semiconductor rings. *Phys. Rev. Lett.* **84**, 2223–2226 (2000).
7. Warburton, R. J. *et al.* Optical emission from a charge-tunable quantum ring. *Nature* **405**, 926–929 (2000).
8. Büttiker, M., Imry, Y. & Landauer, R. Josephson behavior in small normal one-dimensional rings. *Phys. Lett. A* **96**, 365–367 (1983).
9. Lévy, L. P., Dolan, G., Dunsmuir, J. & Bouchiat, H. Magnetization of mesoscopic copper rings: Evidence for persistent currents. *Phys. Rev. Lett.* **64**, 2074–2077 (1990).
10. Chandrasekhar, V. *et al.* Magnetic response of a single, isolated gold loop. *Phys. Rev. Lett.* **67**, 3578–3581 (1991).
11. Mailly, D., Chapelier, C. & Benoit, A. Experimental observation of persistent currents in GaAs–AlGaAs single loop. *Phys. Rev. Lett.* **70**, 2020–2023 (1993).
12. Kouwenhoven, L. P. *et al.* in *Nato ASI Conf. Proc.* (eds Kouwenhoven, L. P., Schön, G. & Sohn, L. L.) 105–214 (Kluwer, Dordrecht, 1997).

13. Tarucha, S., Austing, D. G., Honda, T., van der Haage, R. J. & Kouwenhoven, L. P. Shell filling and spin effects in a few electron quantum dot. *Phys. Rev. Lett.* **77**, 3613–3616 (1996).
14. Held, R. *et al.* In-plane gates and nanostructures fabricated by direct oxidation of semiconductor heterostructures with an atomic force microscope. *Appl. Phys. Lett.* **73**, 262–264 (1998).
15. Lüscher, S., Heinzel, T., Ensslin, K., Wegscheider, W. & Bichler, M. Signatures of spin pairing in a quantum dot in the Coulomb blockade regime. *Phys. Rev. Lett.* **86**, 2118–2121 (2001).
16. Heinzel, T. *et al.* Electronic properties of semiconductor nanostructures patterned by AFM lithography. *Physica E* **9**, 84–93 (2001).
17. Pedersen, S., Hansen, A. E., Kristensen, A., Sørensen, S. B. & Lindelof, P. E. Observation of quantum asymmetry in an Aharonov-Bohm ring. *Phys. Rev. B* **61**, 5457–5460 (2000).
18. Cassé, M. *et al.* Temperature dependence of the Aharonov-Bohm oscillations and the energy spectrum in a single-mode ballistic ring. *Phys. Rev. B* **62**, 2624–2629 (2000).
19. Tan, W.-C. & Inkson, J. C. Electron states in a two-dimensional ring—an exactly soluble model. *Semicond. Sci. Technol.* **11**, 1635–1641 (1996).
20. Chakraborty, T. & Pietiläinen, P. Persistent currents in a quantum ring: Effects of impurities and interactions. *Phys. Rev. B* **52**, 1932–1935 (1995).
21. Kastner, M. A. The single-electron transistor. *Rev. Mod. Phys.* **64**, 849–858 (1992).
22. Berman, D., Entin-Wohlman, O. & Azebel, M. Y. Diamagnetic spectrum and oscillations in an elliptical shell. *Phys. Rev. B* **42**, 9299–9306 (1990).
23. Beenakker, C. W. J. Random-matrix theory of quantum transport. *Rev. Mod. Phys.* **69**, 731–808 (1997).
24. Schuster, R. & Ensslin, K. Antidot superlattices: classical chaos and quantum transport. *Adv. Solid State Phys.* **34**, 195–218 (1994).
25. Loss, D. & Goldbart, P. Period and amplitude halving in mesoscopic rings with spin. *Phys. Rev. B* **43**, 13762–13765 (1991).

Acknowledgements

We thank M. Büttiker and D. Loss for valuable discussions. Financial support from the Swiss Science Foundation (Schweizerischer Nationalfonds) is gratefully acknowledged.

Correspondence and requests for materials should be addressed to K.E. (e-mail: ensslin@phys.ethz.ch).

Coherent control of pulsed X-ray beams

M. F. DeCamp*, D. A. Reis*, P. H. Bucksbaum*, B. Adams†, J. M. Caraher*, R. Clarke*, C. W. S. Conover‡, E. M. Dufresne*, R. Merlin*, V. Stoica* & J. K. Wahlstrand*

* Department of Physics and FOCUS Center, University of Michigan, Ann Arbor, Michigan 48109, USA

† Advanced Photon Source, Argonne National Labs, Argonne, Illinois 60439, USA

‡ Colby College, Waterville, Maine 04901, USA

Synchrotrons produce continuous trains of closely spaced X-ray pulses. Application of such sources to the study of atomic-scale motion requires efficient modulation of these beams on time-scales ranging from nanoseconds to femtoseconds. However, ultrafast X-ray modulators are not generally available. Here we report efficient subnanosecond coherent switching of synchrotron beams by using acoustic pulses in a crystal to modulate the anomalous low-loss transmission of X-ray pulses. The acoustic excitation transfers energy between two X-ray beams in a time shorter than the synchrotron pulse width of about 100 ps. Gigahertz modulation of the diffracted X-rays is also observed. We report different geometric arrangements, such as a switch based on the collision of two counter-propagating acoustic pulses: this doubles the X-ray modulation frequency, and also provides a means of observing a localized transient strain inside an opaque material. We expect that these techniques could be scaled to produce subpicosecond pulses, through laser-generated coherent optical phonon modulation of X-ray diffraction in crystals. Such ultrafast capabilities have been demonstrated thus far only in laser-generated X-ray sources, or through the use of X-ray streak cameras^{1–6}.

X-ray anomalous transmission (the Borrmann effect) is a classical diffraction effect where X-rays propagate through a crystal with low

loss over many average attenuation lengths^{7–9}. The X-ray field found by solving Maxwell's equations in a periodic medium has two linearly independent eigensolutions (see Fig. 1). Solution α is the anomalous transmission wave, where absorption is reduced because the nodes of the X-ray field lie near the atomic crystal planes, as in a waveguide. Solution β has the antinodes of the field near the lattice planes, creating enhanced absorption. In crystals thicker than several attenuation lengths, the β -solution is almost completely absorbed, and the X-rays are nearly pure α -type. At the exit face of the crystal, the α -wave redistributes into two freely propagating X-ray beams with approximately equal intensities: the forward diffracted beam, with a momentum parallel to the original incident beam; and a deflected diffracted beam, which satisfies the Laue condition for a particular set of crystal planes.

Anomalous transmission is sensitive to small lattice distortions. Borrmann's own experiments demonstrated the sensitivity of anomalous transmission to lattice strain due to a small thermal gradient¹⁰. More recent work has shown that low-frequency acoustic waves can spatially modify the anomalous transmission¹¹ or even destroy it^{12–14}. In all of these experiments, the induced strain weakens the anomalous transmission.

The forward and deflected beams are the eigenmodes outside the crystal. At the exit face of the crystal, α and β redistribute into orthogonal linear combinations of these two freely propagating mutually coherent X-ray beams. The α - and β -solutions propagate at different phase velocities, because they experience different indices of refraction. This leads to the Pendellösung effect—that is, the modulation of the intensity of the outgoing X-ray beams as a function of the total accumulated phase difference between the two interior solutions—provided that the crystal is thin enough for the β -solution not to be absorbed^{7–9}. The key to coherent control of the X-ray pulses is to create coherent combinations of α and β near the exit face of the crystal, which will interfere to produce the desired modulation in the free solutions outside the crystal. In particular, a buried interface (such as a dislocation) repopulates the β -solution even after it has decayed away in a thick crystal¹⁵.

In previous experiments on anomalous transmission^{11–14}, the perturbations encompassed the entire crystal bulk. In the present

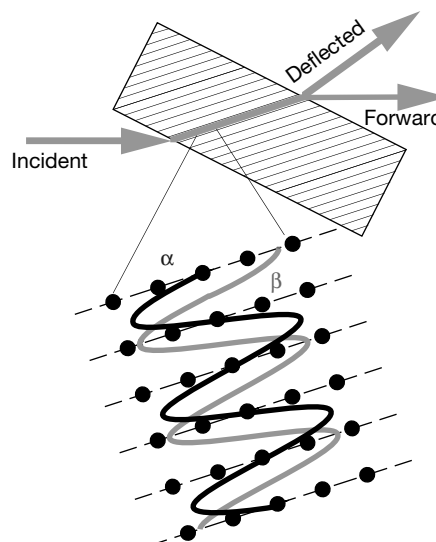


Figure 1 Geometry for X-ray diffraction from the asymmetric $\bar{2}02$ crystal planes of Ge[001]. Inside the crystal, the field consists of two transverse standing waves: the α -wave has its nodes on the atomic planes and thus experiences low absorption, while the β -wave has its antinodes on the atomic planes and thus experiences enhanced absorption. At the exit face of the crystal, α and β redistribute into orthogonal linear combinations of two freely propagating X-ray beams: a forward-diffracted beam and a deflected-diffracted beam.

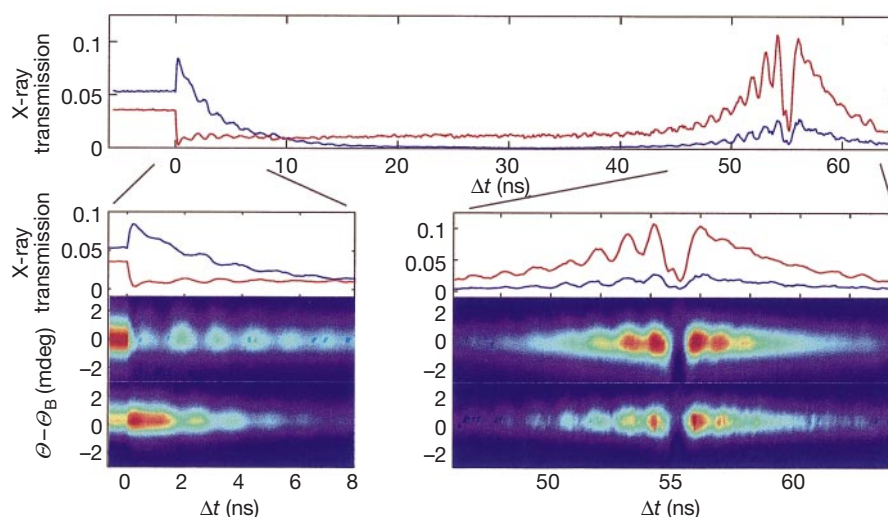


Figure 2 Time-resolved measurement of the $\bar{2}02$ diffraction following laser impulsive excitation of an acoustic wave on the exit face of a thick Ge [001] crystal. Top: forward (blue) and deflected (red) X-ray intensity near the diffraction peak as a function of time delay (Δt) following excitation. Bottom: expanded scale showing the regions where the

acoustic pulse is near the exit or entrance crystal surface. Two-dimensional plots: the measured X-ray transmission versus Δt and crystal angle ($\Theta - \Theta_B$) near the forward (top) and deflected (bottom) diffraction peaks in normalized false colour.

work, an ultrafast (70-fs) laser pulse is absorbed in a 200-nm-thick surface layer of a 280- μm -thick Ge single crystal. This launches an acoustic impulse into the crystal^{16–20}, which modulates the ratio of the two diffracted X-ray beams on a subnanosecond timescale. The laser absorption, and the accompanying acoustic impulse, occurs over a length much less than the characteristic Pendellösung length ($\sim 5\text{ }\mu\text{m}$). The impulse can be modelled as a thin interface that separates two distinct but otherwise perfect crystals. This interface rotates the wave solution in the (α, β) basis (thus maintaining their mutual coherence), in effect modifying the boundary conditions to the second crystal. By adjusting the delay between the X-rays and the acoustic pulse we can cause the exit beams to switch energy preferentially from one beam to another, modulate the two beams together, or even enhance the anomalous transmission.

The experiments were performed at the Michigan-Howard-Lucent Technologies undulator beamline at the Advanced Photon Source synchrotron. The spacing between X-ray pulses was 152 ns, sufficient to detect single X-ray pulses by gating the signal from a silicon avalanche photodiode (APD). Two independent APDs were used to measure the forward and deflected beams simultaneously. A Si(111) double crystal monochromator delivered up to 10^7 10-keV X-ray photons per pulse in a bandwidth of 1.4 eV. The X-ray switch was a 280- μm crystal of Ge[001], polished on both faces, which was impulsively excited by a 70-fs laser pulse, at 840 nm wavelength, with a fluence of $\sim 5\text{ mJ cm}^{-2}$. The X-ray attenuation length at 10 keV is 50 μm . The laser was synchronized to the X-ray pulses with a precision of 19 ps, over a delay of -1 to 1 ms. Our measurement resolution was limited by the X-ray pulse length of ~ 100 ps.

Figure 1 illustrates the transmission geometry for X-rays incident in a direction that satisfies the Bragg condition for the $\bar{2}02$ crystal planes of a [001]-oriented crystal. This geometry is asymmetric, in that the forward and deflected beams leave the crystal at different angles with respect to the surface normal. Inside the crystal, the X-rays propagate at 45° with respect to the surface normal, parallel to the crystal planes, such that the effective propagation distance was about eight average attenuation lengths. Figure 2 shows the time-resolved transmission following ultrafast laser-excitation on the exit face of the crystal (the face at which the α - and β -solutions redistribute into the freely propagating forward and deflected beams). We detected both the forward and the deflected X-rays as a function Δt and as a function of crystal angle, near the diffraction peak. (Δt is the time difference between the arrival of the X-ray pulse

and the laser pulse; the deviation from the Bragg condition is given by $\Theta - \Theta_B$.)

At time delays less than $\Delta t = 0$, the X-rays arrive at the crystal before any laser excitation, and propagate via the Borrmann effect. The incident beam outside the crystal splits into α and β immediately inside the surface, and because the crystal is thick, only the α portion propagates with little attenuation. At the output face, α splits into approximately equal portions of forward and deflected beam.

Starting at $\Delta t = 0$, the diffracted intensity switches from the deflected beam to the forward beam. Approximately 75% of the energy is transferred in less than 100 ps. The amount of energy transfer was observed to depend on the incident laser fluence. As only the α -solution exists at the exit face, a full transfer of energy requires repopulation of the β solution. After X-ray propagation over one-quarter of a Pendellösung period (which is greater than 1.25 μm for 10-keV X-rays in this geometry), the output has switched to pure forward or deflected X-rays, depending on the initial sign of the rotation. Assuming an upper limit of the switch time given by our resolution of 100 ps, this implies that either the Pendellösung length decreases by a factor of at least 2.5 or the excitation propagates into the crystal at greater than 2.5 times the sound speed (5,000 m s^{-1} for longitudinal acoustic phonons along [001]). To resolve the initial switch mechanism, faster time resolution (for example a picosecond X-ray streak camera) may be required.

The Pendellösung length decreases as the crystal tilts away from the Laue resonance. A transient strain with components along the reciprocal lattice vector could reduce the oscillation length enough to switch the beams. Another possible mechanism is the supersonic transfer of energy through the ambipolar diffusion of the dense electron-hole plasma produced by laser excitation. The plasma then provides the source of stress that produces lattice strain in the crystal^{17,21,22}. Estimates based on our experimental parameters indicate that either mechanism is plausible.

Following the initial switch in Fig. 2, Pendellösung oscillations are apparent in both the forward and deflected beams. The oscillations appear to be π radians out of phase with a frequency of ~ 1 GHz. This frequency is consistent with the passage of an acoustic impulse that redistributes energy from the α wave to a linear combination of α and β as it moves along the [001] crystal normal at the speed of sound. These oscillations decay with a time constant ~ 5 ns. This is

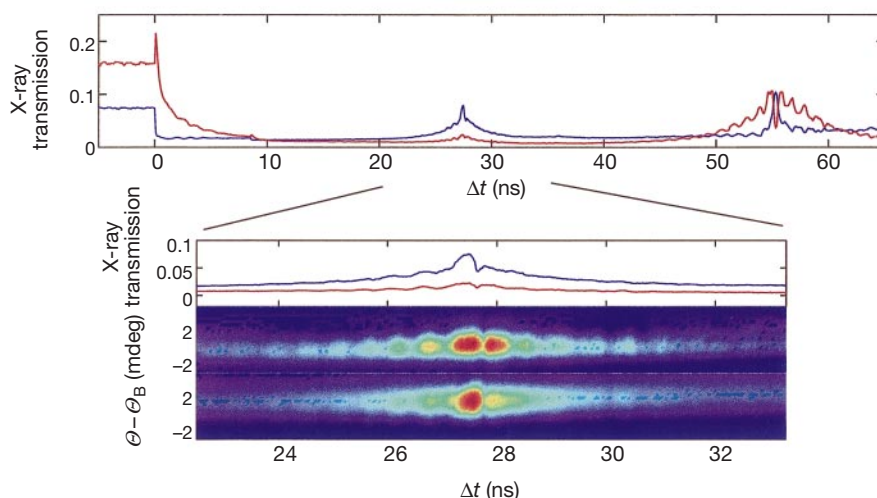


Figure 3 Time-resolved measurement of the $20\bar{2}$ diffraction following simultaneous impulsive excitation of an acoustic wave on the entrance and exit faces of a thick Ge[001] crystal. Top: forward (blue) and deflected (red) X-ray intensity near the diffraction peak as a function of time delay (Δt) following excitation. Bottom: expanded scale showing the

region where the two acoustic pulses cross in the centre of the crystal. Two-dimensional plots: the measured X-ray transmission versus Δt and crystal angle ($\Theta - \Theta_B$) near the forward (top) and deflected (bottom) diffraction peaks in normalized false colour.

faster than an average attenuation length divided by the speed of sound (~ 10 ns), owing to the enhanced absorption of the β solution.

Around 55 ns the acoustic pulse approaches the input face of the crystal at the sound speed, so that the acoustic disturbance forms a thin interface separating the near-input region from the rest of the crystal. The diffraction efficiency increases as the interface approaches the input face because the initial β solution has not yet decayed. Once again, the time constant is determined by the β absorption and the speed of sound. The disturbance is such that the rotation of the (α, β) basis results in an increased amplitude in the α solution. Pendellösung oscillations are apparent as the interface moves through the crystal in this near-surface region. However, the two diffracted beams now oscillate in phase because the β solution is severely attenuated by the following thick crystal; that is, the two external fields are determined solely by the α -solution (anomalous transmission) as it is modified at the interface.

We can double the Pendellösung frequency by colliding two counter-propagating acoustic disturbances. Figure 3 shows the X-ray transmission in the forward and deflected beams following simultaneous excitation of acoustic waves on the input and output crystal faces. In this case the crystal was oriented so that the incident wave diffracts from the $20\bar{2}$ planes. This is similar to the $20\bar{2}$ geometry (Fig. 1) except that the incident beam is now along a direction parallel to the $20\bar{2}$ deflected beam, that is, the opposite asymmetry.

Two acoustic pulses separate the crystal into three regions. Near $\Delta t = 0$ and again near $\Delta t = 55$ ns a thick inner crystal is separated by thinner crystals at the entrance and exit faces. The initial switch increases the intensity of the deflected beam and decreases the intensity in the forward beam. This is in contrast to the effect observed with the switch using the opposite asymmetry. We also observed (not shown) that single-pulse excitation in this geometry results in a similar initial redistribution of energy. We infer that the sign of α/β at the point of the disturbance changes with the crystal orientation asymmetry. Near $\Delta t = 27.5$ ns, or half of the acoustic transit time, a thin crystal develops in the centre of the Ge. This is accompanied by a revival of the Borrmann effect that is particularly evident in the deflected beam. Pendellösung oscillations are again visible but the oscillation frequency is now ~ 2 GHz, indicating that two interfaces are approaching each other at twice the speed of sound. As Δt approaches 55 ns, the oscillations are out of phase

because the β -solution is repopulated by the acoustic pulse launched from the input surface at $\Delta t = 0$.

We have created an efficient X-ray switch that can transfer energy between two widely separated X-ray beams at speeds faster than 100 ps. The motion of atoms in the crystal lattice limits the ultimate speed of this switch. Optical phonon excitations in this geometry could lead to subpicosecond switching times. □

Received 13 June; accepted 3 September 2001.

- Kieffer, J.-C. *et al.* Ultrafast X-ray sources. *Phys. Fluids B* **5**, 2676–2681 (1993).
- Rousse, A. *et al.* Efficient K α x-ray source from femtosecond laser-produced plasmas. *Phys. Rev. E* **50**, 2200–2207 (1994).
- Murnane, M. M. *et al.* Ultrafast X-ray pulses from laser-produced plasmas. *Science* **251**, 531–536 (1991).
- Schoenlein, R. W. *et al.* Femtosecond X-ray pulses at 0.4 Å generated by Thomson scattering: A tool for probing the structural dynamics of materials. *Science* **274**, 236–238 (1996).
- Schoenlein, R. W. *et al.* Generation of femtosecond pulses of synchrotron radiation. *Science* **287**, 2237–2240 (2000).
- Larsson, J. *et al.* Ultrafast x-ray diffraction using a streak-camera detector in averaging mode. *Opt. Lett.* **22**, 1012–1014 (1997).
- Batterman, B. & Cole, H. Dynamical diffraction of X rays by perfect crystals. *Rev. Mod. Phys.* **36**, 681–716 (1964).
- Warren, B. E. *X-Ray Diffraction* (Dover, New York, 1990).
- Zachariasen, W. H. *Theory of X-ray Diffraction in Crystals* (Wiley and Sons, New York, 1945).
- Borrmann, G. & Hildebrandt, G. Röntgen-Wellenfelder in Grossen Kalkspatkristallen und die Wirkung Einer Deformation. *Z. Naturforsch.* **A 11**, 585–587 (1956).
- Hauer, A. & Burns, S. J. Observation of an x-ray shuttering mechanism utilizing acoustic interruption of the Borrmann effect. *Appl. Phys. Lett.* **27**, 524–526 (1975).
- Entin, I. R. *et al.* X-ray acoustic resonance in a perfect silicon crystal. *Sov. Phys. Solid State* **20**, 754–756 (1978).
- LeRoux, S. D., Colella, R. & Bray, R. X-ray diffraction studies of acoustoelectrically amplified phonon beams. *Phys. Rev. Lett.* **35**, 230–234 (1975).
- Chapman, L. D., Colella, R. & Bray, R. X-ray diffraction studies of acoustoelectrically amplified phonons. *Phys. Rev. B* **27**, 2264–2277 (1983).
- Authier, A., Lagomarsino, S. & Tanner, B. K. (eds) *X-ray and Neutron Dynamical Diffraction, Theory and Applications* (Plenum, New York, 1996).
- Thomsen, C. *et al.* Surface generation and detection of phonons by picosecond light pulses. *Phys. Rev. B* **34**, 4129–4138 (1986).
- Akhmanov, S. A. & Gusev, V. E. Laser excitation of ultrashort acoustic pulses: new possibilities in solid-state spectroscopy, diagnostics of fast processes, and nonlinear acoustics. *Sov. Phys. Usp.* **35**, 153–191 (1992).
- Rose-Petruck, C. *et al.* Picosecond-milliangstrom lattice dynamics measured by ultrafast X-ray diffraction. *Nature* **398**, 310–312 (1999).
- Lindenberg, A. *et al.* Time-resolved X-ray diffraction from coherent phonons during a laser induced phase transition. *Phys. Rev. Lett.* **84**, 111–114 (2000).
- Reis, D. A. *et al.* Probing impulsive strain with X-ray pulses. *Phys. Rev. Lett.* **86**, 3072–3075 (2001).
- Auston, D. H. & Shank, C. V. Picosecond ellipsometry of transient electron-hole plasmas in germanium. *Phys. Rev. Lett.* **32**, 1120–1123 (1974).
- Young, J. F. & van Driel, H. M. Ambipolar diffusion of high-density electrons and holes in Ge, Si, and GaAs: many-body effects. *Phys. Rev. B* **26**, 2147–2158 (1982).

Acknowledgements

We thank R. Colella and J. Wark for discussions, and D. Arms for technical support. This work was conducted at the MHATT-CAT insertion device beamline at the Advanced Photon Source. Use of the Advanced Photon Source was supported by the US Department of Energy Basic Energy Sciences, Office of Energy Research. We also acknowledge partial support from the National Science Foundation, the Center for Ultrafast Optical Science, and the AFOSR through the MURI programme.

Correspondence and requests for materials should be addressed to M.F.D.
(e-mail: mattydee@umich.edu).

Spin-dependent exciton formation in π -conjugated compounds

J. S. Wilson*, A. S. Dhoot*, A. J. A. B. Seeley*, M. S. Khan†, A. Köhler* & R. H. Friend*

* Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, UK

† Department of Chemistry, Sultan Qaboos University, Al-Khod 123, Oman

The efficiency of light-emitting diodes (LEDs) made from organic semiconductors is determined by the fraction of injected electrons and holes that recombine to form emissive spin-singlet states rather than non-emissive spin-triplet states. If the process by which these states form is spin-independent, the maximum efficiency of organic LEDs will be limited to 25 per cent¹. But recent reports have indicated fractions of emissive singlet states ranging from 22 to 63 per cent^{2–5}, and the reason for this variation remains unclear. Here we determine the absolute fraction of singlet states generated in a platinum-containing conjugated polymer and its corresponding monomer. The spin-orbit coupling introduced by the platinum atom allows triplet-state emission, so optically and electrically generated luminescence from both singlet and triplet states can be compared directly. We find an average singlet generation fraction of 22 ± 1 per cent for the monomer, but 57 ± 4 per cent for the polymer. This suggests that recombination is spin-independent for the monomer, but that a spin-dependent process, favouring singlet formation, is effective in the polymer. We suggest that this process is a consequence of the exchange interaction, which will operate on overlapping electron and hole wavefunctions on the same polymer chain at their capture radius.

Several sophisticated methods have been used to determine the singlet generation fraction in organic LEDs, but these are often complicated by the fact that triplet states are generally non-emissive in conjugated organic compounds. As a result it has previously only been possible to place a lower limit on the singlet generation fraction in polymer LEDs. We have circumvented the problem of a non-emissive triplet state by investigating a Pt-containing polymer and its corresponding monomer (Fig. 1). The heavy Pt atom introduces spin-orbit coupling which allows the spin of the electron to flip or rephase and transitions between the singlet and triplet manifolds become possible⁶. Conjugation is preserved through the Pt atom by mixing of the frontier orbitals of the Pt and the ligand⁷. Direct emission from the triplet state is therefore observed for both optical and electrical excitation of LEDs made from these materials and this allows a straightforward calculation of the absolute singlet generation fraction.

The photoluminescence and electroluminescence spectra of the polymer and monomer (Fig. 2) both show two characteristic emission bands, and the relative contribution of each band to the total emission is indicated as a percentage. The high and low energy bands are due to singlet and triplet excited states S_1 and T_1 ,

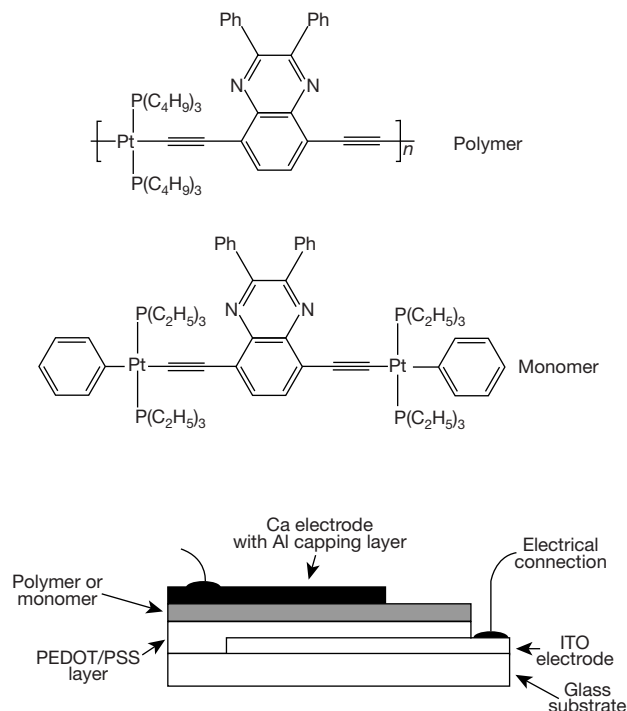


Figure 1 The chemical structures of the platinum-containing polymer and monomer investigated. A schematic of the simple light-emitting diode structure used is also given. PEDOT, poly(3,4-ethylenedioxythiophene); PSS, poly(styrenesulphonate); ITO, indium-tin oxide.

respectively. The triplet-state emission of this polymer and monomer and that of other similar materials has been well characterized previously by lifetime and photoinduced absorption measurements^{8–11}. The triplet emission is observed at around 0.7 eV below the singlet in this and other conjugated polymers^{9,10,12–14} and this finite exchange energy implies an excited state which is self-localized.

We note that although the spin-orbit coupling induced by the heavy Pt atom has been shown to produce close to 100% intersystem crossing⁸ from the S_1 state to T_1 , there is still more emission observed from S_1 than from T_1 in photoluminescence at 290 K. This is because, despite the T_1 – S_0 transition being partially allowed in this polymer and monomer, the radiative decay rates from the triplet states are still only of the order of 10^3 s^{–1} in comparison with non-radiative decay rates of 10^6 – 10^5 s^{–1}, so only one in 1,000 of the triplets generated emits¹⁵. The non-radiative decay rate increases exponentially with decreasing T_1 – S_0 gap and is controlled by multi-phonon emission¹⁵.

In order to compare electroluminescence and photoluminescence spectra we measured both in LED structures. For electrical excitation we observe a greater percentage of the photons to be from the triplet state than with optical excitation, but in other respects the spectra are very similar (Fig. 2). In order to use the emission spectra to calculate the fraction of excited states generated in the singlet spin state, χ_S , we model the processes of photoluminescence and electroluminescence as shown in Fig. 3. In photoluminescence a number I of excitons are originally all created in the singlet S_1 state, and from there they decay radiatively (with a decay rate $k_{R(S)}$) or non-radiatively (with a decay rate $k_{NR(S)}$) to the singlet ground state S_0 , or undergo intersystem crossing to the triplet manifold (with a rate k_{ISC}). Both radiative and non-radiative decay occurs from the triplet state T_1 to the ground state S_0 . For electroluminescence the processes are exactly the same as in photoluminescence, with the important exception that it is also possible to create a fraction χ_T of triplet excitons directly. We define the quantum yields

Φ for radiative triplet (TR) and singlet (SR) emission and for intersystem crossing (ISC):

$$\Phi_{\text{SR}} = \frac{k_{\text{R(S)}}}{[k_{\text{ISC}} + k_{\text{R(S)}} + k_{\text{NR(S)}}]} \quad \Phi_{\text{TR}} = \frac{k_{\text{R(T)}}}{[k_{\text{R(T)}} + k_{\text{NR(T)}}]}$$

$$\Phi_{\text{ISC}} = \frac{k_{\text{ISC}}}{[k_{\text{ISC}} + k_{\text{R(S)}} + k_{\text{NR(S)}}]}$$

The number of photons emitted from the triplet and singlet states are denoted N_{T} and N_{S} respectively, and the ratio of these is labelled R . For photoluminescence:

$$N_{\text{S,PL}} = I\Phi_{\text{SR,PL}}$$

$$N_{\text{T,PL}} = I\Phi_{\text{ISC,PL}}\Phi_{\text{TR,PL}}$$

$$R_{\text{PL}} = \frac{N_{\text{T,PL}}}{N_{\text{S,PL}}} = \frac{\Phi_{\text{ISC,PL}}\Phi_{\text{TR,PL}}}{\Phi_{\text{SR,PL}}}$$

For electroluminescence:

$$N_{\text{S,EL}} = I\chi_{\text{S}}\Phi_{\text{SR,EL}}$$

$$N_{\text{T,EL}} = I\chi_{\text{S}}\Phi_{\text{ISC,EL}}\Phi_{\text{TR,EL}} + I\chi_{\text{T}}\Phi_{\text{TR,EL}}$$

$$R_{\text{EL}} = \frac{N_{\text{T,EL}}}{N_{\text{S,EL}}} = \frac{\chi_{\text{S}}\Phi_{\text{ISC,EL}}\Phi_{\text{TR,EL}} + \chi_{\text{T}}\Phi_{\text{TR,EL}}}{\chi_{\text{S}}\Phi_{\text{SR,EL}}}$$

Comparing photoluminescence and electroluminescence gives:

$$\frac{R_{\text{PL}}}{R_{\text{EL}}} = \left(\frac{\chi_{\text{S}}\Phi_{\text{ISC,PL}}}{\chi_{\text{S}}\Phi_{\text{ISC,EL}} + \chi_{\text{T}}} \right) \left(\frac{\Phi_{\text{TR,PL}}}{\Phi_{\text{TR,EL}}} \right) \left(\frac{\Phi_{\text{SR,EL}}}{\Phi_{\text{SR,PL}}} \right) \quad (1)$$

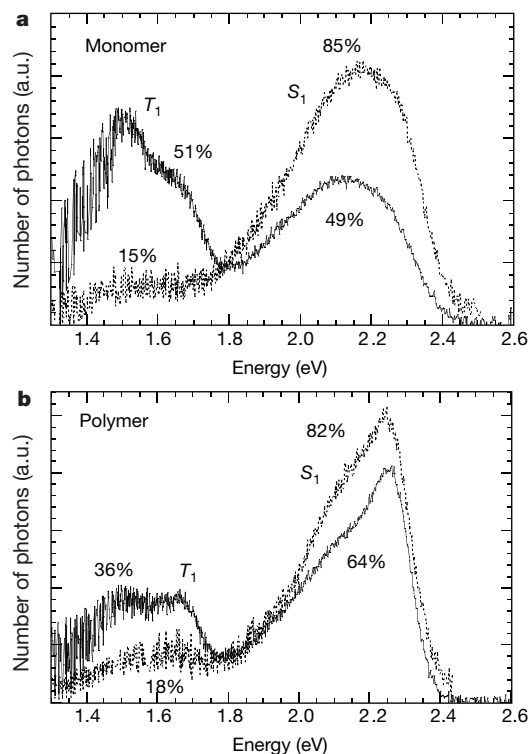


Figure 2 Comparison of the photoluminescence and electroluminescence spectra of light-emitting diodes of the platinum-containing polymer and monomer at 290 K. **a**, The photoluminescence (dotted line) and electroluminescence (solid line) spectra of a monomer LED. **b**, The photoluminescence and electroluminescence spectra of a polymer LED. The triplet emission is denoted by T_1 and the singlet emission by S_1 . The spectra have been plotted in terms of number of photons and the fraction of the total emission that the singlet and triplet emission represent are given as percentages.

If we assume that the decay rates are the same for electroluminescence and photoluminescence we obtain $\Phi_{\text{PL}} = \Phi_{\text{EL}}$, and because $\chi_{\text{S}} + \chi_{\text{T}} = 1$, equation (1) becomes:

$$\frac{R_{\text{PL}}}{R_{\text{EL}}} = \left(\frac{\chi_{\text{S}}\Phi_{\text{ISC,PL}}}{\chi_{\text{S}}\Phi_{\text{ISC,EL}} + (1 - \chi_{\text{S}})} \right) \quad (2)$$

For these materials Φ_{ISC} is known to be close to one⁸ so we approximate equation (2) by:

$$\frac{R_{\text{PL}}}{R_{\text{EL}}} = \chi_{\text{S}} \quad (3)$$

Therefore, simply by dividing the spectrally integrated triplet emission by that of singlet emission for electroluminescence and photoluminescence in spectra such as those in Fig. 2, it is possible to obtain an absolute value for the singlet generation fraction, χ_{S} , with the advantage that many factors involved cancel out. For example for the data shown in Fig. 2b, $R_{\text{PL}} = 18/82$, and $R_{\text{EL}} = 36/64$ which gives $\chi_{\text{S}} = 0.39$.

The only assumption that has been made is that there is no difference in the decay rates for excited states whether they are created optically or electrically and we now examine this more closely. If electroluminescence and photoluminescence originate from different positions in the device relative to the electrodes, they may experience different quenching^{16,17}. To check this we measured the dependency of χ_{S} on film thickness (Fig. 4a) and also the lifetimes (and therefore decay rates⁶) of triplet emission for both optical and electrical excitation. We find no significant variations in χ_{S} down to devices where the emitting layer was only around 20 nm thick and we observe the same lifetimes in photoluminescence and electroluminescence. For the monomer the room-temperature triplet lifetime is $10 \pm 2 \mu\text{s}$ for optical excitation and $11.8 \pm 0.5 \mu\text{s}$ for electrical excitation. For the polymer the room temperature lifetime is shorter and therefore more difficult to measure, but at around 100 K optical excitation gives a triplet lifetime of $2.8 \pm 0.8 \mu\text{s}$ and electrical excitation gives $3 \pm 1 \mu\text{s}$. The decay rates of the triplet are therefore the same for both optical and electrical excitation.

In order to elucidate which parameters control the generation of singlet and triplet excited states in organic LEDs we studied the dependence of the singlet generation fraction on electric field and temperature. The temperature may affect carrier mobilities and processes with an activation energy, and the electric field could also influence exciton formation¹⁸. However, it can be seen from Fig. 4b and c that we observe no significant variation in χ_{S} with temperature or electric field over the working range of the devices. We note that the room-temperature values for χ_{S} are slightly lower than the values below 200 K, but we cannot rule out a systematic error here as

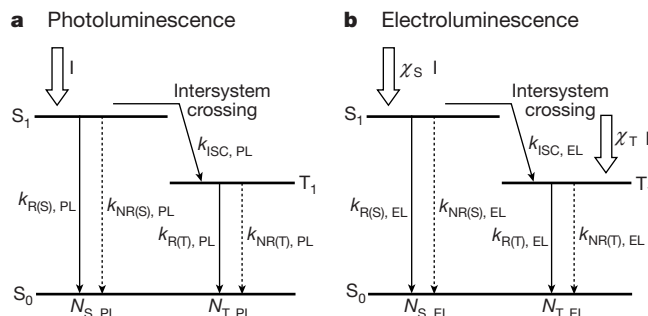


Figure 3 A simple model for the emission spectra in photoluminescence (**a**) and electroluminescence (**b**). I represents an initial number of excitons. k_{NR} , k_{R} and k_{ISC} represent non-radiative, radiative and intersystem crossing rates respectively, and N_{S} and N_{T} represent the number of singlet and triplet photons that would be expected to be emitted. χ_{S} and χ_{T} are the singlet and triplet generation fractions, respectively. S_0 is the singlet ground state.

there is very little triplet emission in photoluminescence at 290 K.

To summarize, we find different values of the singlet generation fraction χ_S for the polymer and monomer of the same material that reconcile the different values available in the literature. The average values of χ_S we obtain for the monomer and polymer at room temperature are 0.15 ± 0.01 and 0.40 ± 0.03 respectively, and averaging over all temperatures gives 0.22 ± 0.01 and 0.57 ± 0.04 respectively. For the monomer this is close to the 0.25 expected from simple spin statistics and in agreement with the value of 0.22 ± 0.03 measured² for the small molecule Alq₃. For the polymer the average value of χ_S is more than double the value expected from simple spin statistics and is above the lower limit for the singlet generation fraction of 0.35–0.5 measured for PPV derivative polymers^{3,4}. The 290 K value also agrees with the χ_S value of 0.4 predicted⁵ for a polymer with a bandgap between 2.3 and 2.5 eV. Clearly the fact that we observe χ_S to be largely independent of temperature and applied electric field means these parameters do not significantly influence the probability of forming either a singlet or triplet spin state.

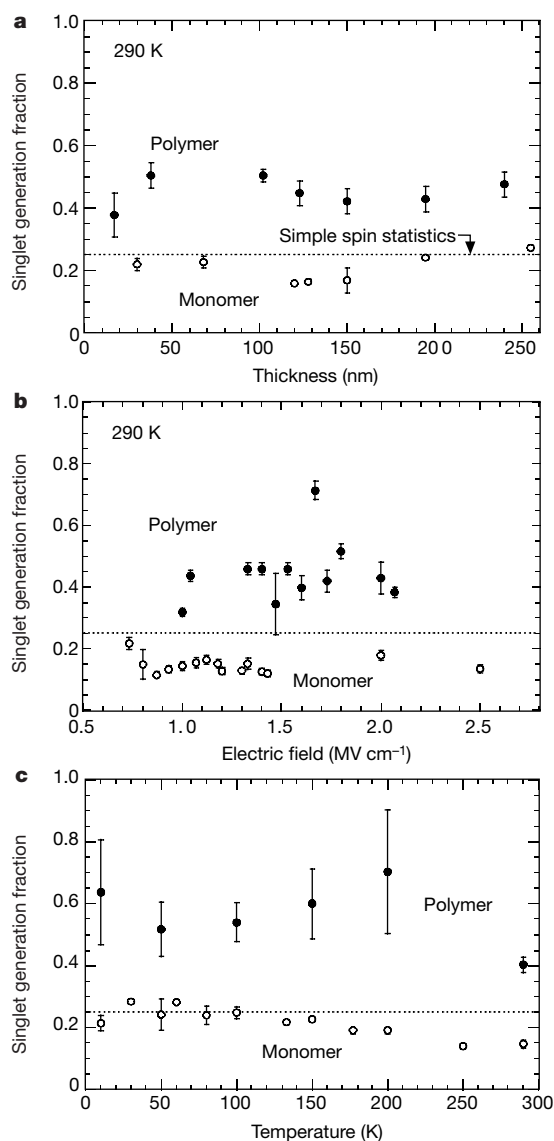


Figure 4 The singlet generation fraction χ_S for the platinum-containing polymer and monomer under different device conditions. **a**, χ_S plotted as a function of device thickness for polymer (solid circles) and monomer (open circles) LEDs. **b**, χ_S plotted as a function of applied electric field. **c**, χ_S plotted as a function of temperature. The error bars on each point represent the statistical error obtained by measuring a large number of devices. The dotted line represents the fraction (0.25) expected from simple spin statistics.

The marked difference we find in χ_S for polymer and monomer LEDs implies different processes of singlet and triplet formation via electron–hole capture and we now consider why these may occur. Organic LEDs operate by injection of electrons and holes from opposite electrodes, and transport by hopping from one molecular site (or polymer chain) to another. Electron and hole capture one another when they are held within their mutually attractive potential, and this process is usually described using the Langevin model^{19,20}. Typical capture radii are about 10 nm. The distinction between molecular semiconductors and conjugated polymers is that this distance is considerably greater than a molecular dimension, but comparable to the delocalization lengths for electron and hole wavefunctions on polymer chains. Therefore, for molecular materials, electron–hole capture is achieved at a range where only the Coulomb interaction is effective, and the process is independent of the final spin state of the exciton (even though the rates for the final step of exciton formation may be spin-dependent). In contrast, for the polymers, when electron and hole are present on the same chain at the capture range, the interaction between them results also from direct overlap of electron and hole wavefunctions, and is therefore spin-dependent (through the exchange interaction). Thus, the probability of electron–hole capture is now spin-dependent. Our results demonstrate that singlets are favoured over triplets, and there are a number of processes which might be responsible for this^{5,18,21–23}. These include the greater spatial extension along the polymer chain of the singlet exciton, giving a stronger longer-range attraction¹⁴; or the closer energy match between the electron–hole and the singlet exciton compared to the triplet exciton¹⁸. Spin-dependent interactions at this point of electron–hole capture give spin-dependent singlet–triplet generation fractions because the probability of forming the exciton over electron and hole moving apart is spin-dependent.

Thus, direct observation of triplet emission has allowed us to demonstrate that there is a different quantum-mechanical process of electron–hole capture in operation in organic compounds with an extended conjugated system as compared to small conjugated molecules. In particular, the process of electron–hole capture in conjugated polymers is dependent on the final spin state. The high singlet generation fractions found in conjugated polymers^{3–5} suggest that it may not be necessary to harvest triplets in these materials in order to create efficient LEDs. □

Methods

Single-layer light-emitting diodes of the platinum-containing polymer and monomer were fabricated according to the structure shown in Fig. 1 by spin-coating and subsequent thermal evaporation of the metal electrodes. Measurements of photoluminescence and electroluminescence were made with the LEDs in a continuous-flow helium cryostat. For photoluminescence, excitation was provided by the 457-nm line of a continuous wave argon ion laser, incident almost perpendicular to the indium–tin oxide (ITO) side of the devices. Measurements of electroluminescence spectra were made with the devices in exactly the same configuration as for photoluminescence and with a constant voltage supplied by a Keithley 230 voltage source. The photoluminescence and electroluminescence emission spectra were recorded perpendicular to the devices by using a spectrograph with an optical fibre input coupled to a cooled charge-coupled device (CCD) array (Oriol Instaspec IV). For the photoluminescence lifetime measurements, the tripled output from a Q-switched YAG laser was used (355 nm, ~15 ns pulses). For the electroluminescence lifetime measurements, square wave electrical pulses (15–30 V, 500 μ s pulses) were applied to the devices. In both cases the transient emission was recorded by using a photomultiplier tube and a digital oscilloscope. Lifetimes were calculated by fitting the decay curves to an exponential. All measurements were made under vacuum or in a He atmosphere of 10 mbar.

Received 5 July; accepted 28 August 2001.

1. Friend, R. H. *et al.* Electroluminescence in conjugated polymers. *Nature* **397**, 121–128 (1999).
2. Baldo, M. A., O'Brien, D. F., Thompson, M. E. & Forrest, S. R. Excitonic singlet–triplet ratio in a semiconducting organic thin film. *Phys. Rev. B* **60**, 14422–14428 (1999).
3. Ho, P. K. H. *et al.* Molecular-scale interface engineering for polymer light-emitting diodes. *Nature* **404**, 481–484 (2000).
4. Cao, Y., Parker, I. D., Yu, G., Zhang, C. & Heeger, A. J. Improved quantum efficiency for electroluminescence in semiconducting polymers. *Nature* **397**, 414–417 (1999).
5. Wohlgenannt, M., Tandon, K., Mazumdar, S., Ramasesha, S. & Vardeny, Z. V. Formation cross-sections of singlet and triplet excitons in π -conjugated polymers. *Nature* **409**, 494–497 (2001).

6. Turro, N. J. *Modern Molecular Photochemistry* 1–195 (University Science, California, 1991).
7. Beljonne, D. *et al.* Spatial extent of the singlet and triplet excitons in transition metal-containing polyynes. *J. Chem. Phys.* **105**, 3868–3877 (1996).
8. Wittmann, H. F., Friend, R. H., Khan, M. S. & Lewis, J. Optical spectroscopy of platinum and palladium containing polyynes. *J. Chem. Phys.* **101**, 2693–2698 (1994).
9. Wilson, J. S. *et al.* Triplet states in a series of Pt-containing ethynyls. *J. Chem. Phys.* **113**, 7627–7634 (2000).
10. Chawdhury, N. *et al.* Evolution of lowest singlet and triplet excited states with number of thienyl rings in platinum polyynes. *J. Chem. Phys.* **110**, 4963–4970 (1999).
11. Chawdhury, N. *et al.* Synthesis and electronic structure of platinum-containing polyynes with aromatic and heteroaromatic rings. *Macromolecules* **31**, 722–727 (1998).
12. Romanovskii, Y. V. *et al.* Phosphorescence of π -conjugated oligomers and polymers. *Phys. Rev. Lett.* **84**, 1027–1030 (2000).
13. Hertel, D. *et al.* Phosphorescence in conjugated poly(para-phenylene)-derivatives. *Adv. Mater.* **13**, 65–70 (2001).
14. Beljonne, D., Shuai, Z., Friend, R. H. & Brédas, J. L. Theoretical investigation of the lowest singlet and triplet states in poly(paraphenylene vinylene) oligomers. *J. Chem. Phys.* **102**, 2042–2049 (1995).
15. Wilson, J. S. *et al.* The energy gap law for triplet states in Pt-containing conjugated polymers and monomers. *J. Am. Chem. Soc.* **123**, 9412–9417 (2001).
16. Becker, H., Burns, S. E. & Friend, R. H. Effect of metal films on the photoluminescence and electroluminescence of conjugated polymers. *Phys. Rev. B* **56**, 1893–1905 (1997).
17. Kim, J. S., Ho, P. K. H., Greenham, N. C. & Friend, R. H. Electroluminescence emission pattern of organic light-emitting diodes: Implications for device efficiency calculations. *J. Appl. Phys.* **88**, 1073–1081 (2000).
18. Kobrak, M. N. & Bittner, E. R. Quantum molecular dynamics study of polaron recombination in conjugated polymers. *Phys. Rev. B* **62**, 11473–11486 (2000).
19. Blom, P. W., de Jong, M. J. M. & Breedijk, S. Temperature dependent electron-hole recombination in polymer light emitting diodes. *Appl. Phys. Lett.* **71**, 930–932 (1997).
20. Pope, M. & Swenberg, C. E. *Electronic Processes in Organic Crystals and Polymers* 2nd edn, 958–960 (Oxford Science, Oxford, 1999).
21. Shuai, Z., Beljonne, D., Silbey, R. J. & Brédas, J. L. Singlet and triplet exciton formation rates in conjugated polymer light-emitting diodes. *Phys. Rev. Lett.* **84**, 131–134 (2000).
22. Hong, T. & Meng, H. Spin-dependent recombination and electroluminescence quantum yield in conjugated polymers. *Phys. Rev. B* **63**, 1–5 (2001).
23. Burin, A. L. & Ratner, M. A. Spin effects on the luminescence yield of organic light emitting diodes. *J. Chem. Phys.* **109**, 6092–6102 (1998).

Acknowledgements

We thank N. C. Greenham for discussions and M. R. A. Al-Mandhary for the synthesis of the polymer and monomer. A.K. thanks Peterhouse, University of Cambridge, UK, for a Fellowship and the Royal Society for a University Research Fellowship. M.S.K. thanks Sultan Qaboos University, Oman, for a research grant and research leave, and the EPSRC (UK) for a Visiting Fellowship. This work was funded by the EPSRC.

Correspondence and requests for materials should be addressed to A.K. (e-mail: ak10007@cam.ac.uk).

Superconductivity in single crystals of the fullerene C₇₀

J. H. Schön*†, Ch. Kloc*, T. Siegrist*‡, M. Steigerwald‡, C. Svensson§ & B. Batlogg*||

* Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

† Department of Physics, University of Konstanz, D-78457 Konstanz, Germany

‡ Agere Systems, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

§ Inorganic Chemistry 2, PO Box 124, Lund University, S-221 00 Lund, Sweden

|| Solid State Physics Laboratory ETH, CH-8093 Zürich, Switzerland

The observation of superconductivity in doped C₆₀ has attracted much attention^{1–3}, as these materials represent an entirely new class of superconductors. A maximum transition temperature (T_c) of 40 K has been reported⁴ for electron-doped C₆₀ crystals, while a T_c of 52 K has been seen⁵ in hole-doped crystals; only the copper oxide superconductors have higher transition temperatures. The results for C₆₀ raise the intriguing questions of whether conventional electron–phonon coupling alone¹ can produce such high transition temperatures, and whether even higher transition temperatures might be observed in other fullerenes^{6–8}. There have, however, been no confirmed reports of superconductivity

in other fullerenes, though it has recently been observed in carbon nanotubes⁹. Here we report the observation of superconductivity in single crystals of electric-field-doped C₇₀. The maximum transition temperature of about 7 K is achieved when the sample is doped to approximately four electrons per C₇₀ molecule, which corresponds to a half-filled conduction band. We anticipate superconductivity in smaller fullerenes at temperatures even higher than in C₆₀ if the right charge density can be induced.

Metallic behaviour is expected for KC₇₀ and K₄C₇₀ owing to the half-filling of the lowest unoccupied molecular orbital bands in K-intercalated C₇₀. An initial transport study on K-intercalated C₇₀ found a maximum conductivity of $\sim 2.5 \text{ cm}^{-1}$ (ref. 10). Theory and experiment showed that K₄C₇₀ exhibits a larger conductivity because of its higher density of states^{11–13}. Pure C₇₀ and alkali-metal-doped C₇₀ show non-zero resistance down to 1.8 K. The absence of superconductivity in these experiments is not well understood, and might be related to disorder and structural defects in the K₄C₇₀ thin films. A weak electron–phonon interaction^{11,14} and a reduced density of states at the Fermi level¹⁵ have also been invoked as reasons for the apparent lack of superconductivity. We note, however, that most experiments have been performed on polycrystalline material, and phase-pure alkali-doped C₇₀ analogues of A₃C₆₀ have not been reported. In addition, various crystallographic structures of C₇₀ have been observed¹⁶. Here we have chosen a different approach to induce electric charge in C₇₀ crystals. We have recently demonstrated that the electrical properties of organic materials can be investigated over a wide range of charge density using field-effect doping, without inducing additional disorder^{5,17–19}, leading to gate-induced superconductivity in a variety of molecular organic crystals^{5,17,18}. This method therefore seems to be well suited to our investigation of the reasons for the absence of superconductivity in the higher fullerenes—is it related to intrinsic properties, or to difficulties in chemical doping? Here we apply this technique to investigate the possibility of superconductivity in high-quality C₇₀ single crystals.

We grew single crystals of C₇₀ (smaller than 1 mm³) from the vapour phase in a stream of hydrogen; the experimental conditions were similar to those used to grow C₆₀ (ref. 5). The crystals were annealed in a xenon atmosphere, but we detected no incorporation of the inert gas into the crystals. We performed an X-ray structural analysis of the C₇₀ single crystals, which showed them to be hexagonal closed packed (h.c.p.) C₇₀, with lattice parameters $a = 10.602(9) \text{ Å}$ and $c = 17.263(13) \text{ Å}$. Previous studies of hexagonal C₇₀ gave lattice parameters of $a = 10.104 \text{ Å}$ and $c = 18.584 \text{ Å}$, with a wide mosaic spread of the order of a few degrees²⁰—furthermore, these crystals showed diffuse rods typical of stacking faults²¹. In contrast, our crystals show a much narrower mosaic of the order of 0.4°. In addition, the c/a ratio for our crystals is close to the ideal value of 1.63 for sphere packing, in contrast to the reported value of 1.84. To first order, it is possible to approximate the ellipsoidal C₇₀ molecule by a hollow sphere with radius $r = 3.78 \text{ Å}$, and we obtained a residual of $R = 0.065$ for 118 independent reflections measured. As stacking faults were expected to be present, we looked for diffuse rods of intensity. Instead, we found well developed superstructure reflections with a mosaic spread of the order of 0.4° consistent with 6H polytypes, leading to a tripling of the c axis ($3c = 51.79 \text{ Å}$). Different 6H stacking sequences are expected, producing twinning at room temperature. On average, five layers are stacked coherently, with the sixth in either the h.c.p. or the cubic close packed (c.c.p.) position. Low-temperature X-ray diffraction revealed a further lowering of symmetry, but the crystals clearly undergo multiple twinning upon cooling. This lowering of the symmetry is indicative of a transition to a phase in which the C₇₀ molecules no longer rotate freely. However, we do not expect that the orientational order of the molecules will be complete, similar to the phase transition observed for C₆₀ upon cooling²². Further structural studies at low temperature are under way.

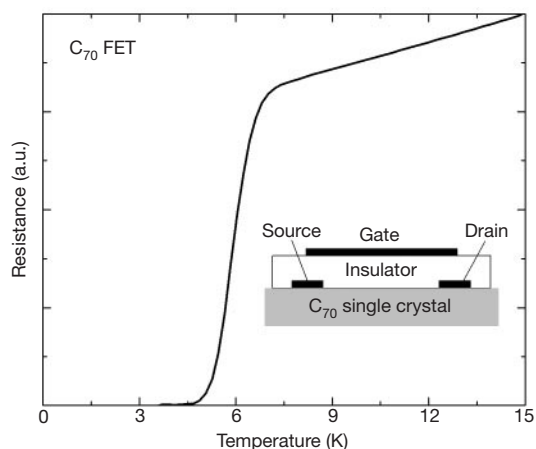


Figure 1 Resistance of a field-effect-doped C_{70} single crystal as a function of temperature. For a doping level of four electrons per molecule, a transition to superconductivity is observed, starting at ~ 7 K. Inset, diagram of the field-effect-device structure.

We prepared field-effect transistors (FETs) on growth surfaces by the deposition of source and drain electrodes made of gold, an Al_2O_3 gate insulator, and a gold gate electrode (see Fig. 1 inset). The transistors showed n-channel activity, with room-temperature mobilities in the range $1\text{--}1.8\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$. Similar mobilities have been reported for C_{60} single-crystal transistors¹⁷. We observed no p-channel activity, a fact that we ascribe to significantly stronger trapping of holes than of electrons. Such behaviour is often observed in organic FETs. We have studied the transport properties of C_{70} p-channel FETs down to 1.7 K in magnetic fields up to 9 T. The experimental field-effect-doping range is limited by the electrical breakdown of the gate insulator.

By applying a strong electric field to the gate, it is possible to induce a very high carrier concentration into the topmost layer of molecules in the organic crystal. Figure 1 shows the channel resistance as a function of temperature for a C_{70} FET. The electron density is approximately four electrons per C_{70} molecule, assuming that only the topmost layer takes part in the conduction^{5,17–19}. The resistance starts to drop at 7 K, reaches its half-point at 6 K, and indicates a transition to the superconducting state. An external magnetic field, applied perpendicular to the FET channel, suppresses the transition as in C_{60} crystals^{5,17}. From the estimated zero-temperature upper critical field, we calculate a coherence length of 40 Å for C_{70} , which is similar to values obtained for bulk-doped as well as field-effect-doped C_{60} (refs 2, 3, 17). The dependence of the transition temperature on the electron density is shown in Fig. 2. Superconductivity (above 1.7 K) is observed for more than three electrons per molecule. A maximum transition temperature of 7 K is observed for four electrons per C_{70} molecule. This density corresponds to half-filling of the conduction band¹². We note that the highest conductivity of alkali-metal-doped C_{70} films has also been observed for this doping level²⁰. However, band structure calculations²⁴ and photoemission data²⁵ have been interpreted as suggesting that the conduction band of C_{70} might take only six electrons. This would then raise the question of why superconductivity in C_{60} occurs around half-filling, and in C_{70} around two-thirds filling. Differences might arise from the cubic and hexagonal crystal structure of C_{60} and C_{70} , respectively. Further detailed theoretical and experimental studies into the electronic structure of C_{70} seem to be appropriate.

It is interesting to compare the transition temperatures of field-effect-doped C_{60} and C_{70} . In the larger molecule T_c is ~ 7 K, whereas it is ~ 11 K in the smaller one. This reduction is in line with the

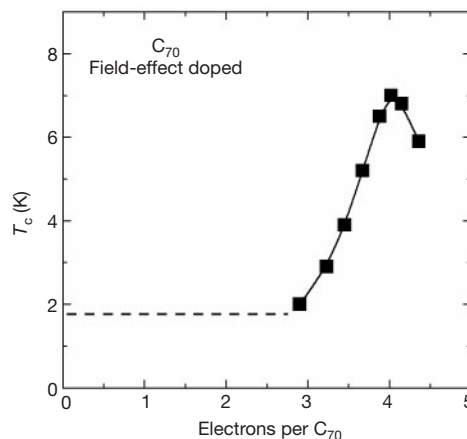


Figure 2 Transition temperature as a function of the doping level. A maximum T_c of 7 K is observed for four electrons per C_{70} molecule. The experimental doping range is limited by the breakdown of the gate insulator. The dashed line indicates the doping range where no superconductivity was observed above 1.7 K.

theoretical expectations of a reduced intra-molecule electron–phonon coupling strength⁶. But other effects also need to be considered—in particular, the electronic density of states and the effective Coulomb interaction strength.

Reports of superconductivity in members of the fullerene family other than C_{60} have not so far been confirmed²³. The present results now indicate the strong possibility that other fullerenes might also become superconducting. This would be another way towards superconductivity at much higher temperatures: the electron–phonon coupling has been calculated to become stronger with decreasing numbers of π -electrons⁶ and with increasing curvature of the molecules^{26,27}. Thus it will be interesting to study the smaller fullerenes^{7,8}, and field-effect doping would seem to be the method of choice to introduce holes and electrons. The present results seem also to indicate the importance of reducing disorder in the crystals or in films. □

Received 26 June; accepted 27 August 2001.

- Hebard, A. F. *et al.* Superconductivity at 18 K in potassium-doped C_{60} . *Nature* **350**, 600–601 (1991).
- Ramirez, A. P. C_{60} and its superconductivity. *Supercond. Rev.* **1**, 1–101 (1994).
- Gunnarsson, O. Superconductivity in fullerenes. *Rev. Mod. Phys.* **69**, 575–606 (1997).
- Palstra, T. T. M. *et al.* Superconductivity at 40 K in cesium doped C_{60} . *Solid State Commun.* **93**, 327–330 (1995).
- Schön, J. H., Kloc, Ch. & Batlogg, B. Superconductivity at 52 K in hole-doped C_{60} . *Nature* **408**, 549–552 (2000).
- Devos, A. & Lannoo, M. Electron-phonon coupling for aromatic molecular crystals: Possible consequences for their superconductivity. *Phys. Rev. B* **58**, 8236–8239 (1998).
- Cohen, M. L. Possibility of high temperature superconductivity in C_{36} and $C_{24}N_{12}$ solids. *Am. Inst. Phys. Conf. Proc.* **483**, 359–365 (1999).
- Piskoti, C., Yarger, J. & Zettl, A. C_{36} , a new carbon solid. *Nature* **393**, 771–774 (1998).
- Tang, Z. K. *et al.* Superconductivity in 4 Å single-walled carbon nanotubes. *Science* **292**, 2462–2465 (2001).
- Haddon, R. C. *et al.* Conducting films of C_{60} and C_{70} by alkali-metal doping. *Nature* **350**, 320–322 (1991).
- Wang, Z. H., Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. Raman studies of electron-phonon interaction in K_3C_{70} . *Phys. Rev. B* **48**, 16881–16884 (1994).
- Saito, S. & Oshiyama, A. Electronic and geometric structures of C_{70} . *Phys. Rev. B* **44**, 11532–11535 (1991).
- Wang, Z. H. *et al.* Electronic properties of K_3C_{70} thin films. *Phys. Rev. B* **48**, 10657–10660 (1994).
- Wang, Z. H., Dresselhaus, M. S., Dresselhaus, G., Wang, K. A. & Eklund, P. C. Electronic properties of K_3C_{70} thin films. II. *Phys. Rev. B* **49**, 15890–15900 (1994).
- Takahashi, T. *et al.* Comparative photoemission study of Rb_3C_{60} , Rb_3C_{70} and Rb_3C_{84} . A pseudo-gap at the Fermi level in the fulleride. *Physica C* **190**, 205–209 (1992).
- Chen, X. H. *et al.* Synthesis, structure, and transport properties of novel fullerenes A_3C_{70} ($A = \text{Ba}$ and Sm). *J. Am. Chem. Soc.* **122**, 5729–5732 (2000).
- Schön, J. H., Kloc, Ch., Haddon, R. C. & Batlogg, B. A superconducting field-effect switch. *Science* **288**, 656–658 (2000).
- Schön, J. H., Kloc, Ch. & Batlogg, B. Superconductivity in molecular crystals induced by charge injection. *Nature* **406**, 704–706 (2000).

19. Schön, J. H. *et al.* Gate-induced superconductivity in solution-processed, organic polymer films. *Nature* **410**, 189–192 (2001).
20. Blanc, H.-B. *et al.* Single-crystal X-ray diffraction study of the room temperature structure and orientational disorder of C_{70} . *Europhys. Lett.* **27**, 359–364 (1994).
21. Blanc, E., Bürgi, H.-B., Restori, R., Schwarzenbach, D. & Ochsenbein, Ph. X-ray diffraction study of the stacking faults in hexagonal C_{70} single crystals. *Europhys. Lett.* **33**, 205–210 (1996).
22. Bürgi, H. B. *et al.* The structure of C_{60} : orientational disorder in the low-temperature modification of C_{60} . *Angew. Chem. Int. Edn Engl.* **31**, 640–643 (1992).
23. Wang, Z. H. *et al.* Electronic properties of K_3C_{70} thin films. *Phys. Rev. B* **48**, 10657–10660 (1994).
24. Ohno, K., Jing-Zhi, Y., Maruyama, Y., Kawazoe, Y. & Takahashi, T. Electronic structures of C_{70} crystalline phases. *Chem. Phys. Lett.* **255**, 367–372 (1996).
25. Golden, M. S. *et al.* The electronic structure of fullerenes and fullerene compounds from high-energy spectroscopy. *J. Phys. Condens. Matter* **7**, 8219–8247 (1995).
26. Kobayashi, M. *et al.* Structure sequence and possible superconductivity in potassium-doped fullerene $C_{70}K_3$. *Phys. Rev. B* **48**, 16877–16880 (1993).
27. Schluter, M., Lannoo, M., Needles, M. & Baraff, G. A. Electron-phonon coupling and superconductivity in alkali-intercalated C_{60} solid. *Phys. Rev. Lett.* **68**, 526–529 (1992).

Acknowledgements

Access to beamline I711 at the MAX-II synchrotron storage ring in Lund is acknowledged.

Correspondence and requests for materials should be addressed to B.B. (e-mail: batlogg@solid.phys.ethz.ch).

Magma storage beneath Axial volcano on the Juan de Fuca mid-ocean ridge

M. West*†, W. Menke*†, M. Tolstoy*, S. Webb*† & R. Sohn‡

* Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

† Department of Earth and Environmental Science, Columbia University, Palisades, New York 10964, USA

‡ Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

Axial volcano, which is located near the intersection of the Juan de Fuca ridge and the Cobb–Eickelberg seamount chain beneath the northeast Pacific Ocean, is a locus of volcanic activity thought to be associated with the Cobb hotspot¹. The volcano rises 700 metres above the ridge, has substantial rift zones extending about 50 kilometres to the north and south, and has erupted as recently as 1998 (ref. 2). Here we present seismological data that constrain the three-dimensional velocity structure beneath the volcano. We image a large low-velocity zone in the crust, consisting of a shallow magma chamber and a more diffuse reservoir in the lower crust, and estimate the total magma volume in the system to be between 5 and 21 km³. This volume is two orders of magnitude larger than the amount of melt emplaced during the most recent eruption^{3,4} (0.1–0.2 km³). We therefore infer that such volcanic events remove only a small portion of the reservoir that they tap, which must accordingly be long-lived compared to the eruption cycle. On the basis of magma flux estimates, we estimate the crustal residence time of melt in the volcanic system to be a few hundred to a few thousand years.

Axial volcano is formed by excess magmatism associated with mantle melting at the Cobb hotspot and the Juan de Fuca ridge⁵. Shoaling bathymetry, inferred crustal thickening⁶, frequent seismic swarms^{7,8} and deformation events^{3,4} are consistent with a robust magma supply system. Lateral dyke injection, observed during a 1998 eruption⁹, originated from a source under the caldera and carried magma up to 50 km along the volcano's rift zones, which are substantial constructional features. Such events suggest that magma supply is strongly focused beneath the volcano. The existence of the 3 × 8 km caldera, its 3 m subsidence during the 1998 eruption¹⁰, and evidence for low-density rock beneath the summit¹¹, suggest a large

magma body a few kilometres beneath the volcano. What has not been understood is how much melt exists, how long it spends in the crust, and how eruptions along the rifts and adjacent ridge segments are related to the magma reservoir. These questions have remained unanswered, not just at Axial volcano, but at basaltic shield volcanoes worldwide.

The compressional-velocity image of the volcano presented here is based on data collected during a 1999 active-source seismic experiment (Fig. 1). 5,025 shots from the airgun array (20 guns totalling 142 litres) of the RV *Maurice Ewing* were recorded on six ocean-bottom seismometers¹² deployed on the volcano's flanks. Water wave travel times and global positioning system (GPS)-determined shot locations were used to precisely locate the seismometers on the sea floor (error < 20 m) and to determine clock drifts (error < 0.02 s). Travel times from 16,400 Pg phases, which turn within the crust, are used in this study.

To characterize the regional structure, a bathymetry-draped one-dimensional (1D) velocity model was assembled from previous studies^{13,14} on the Juan de Fuca ridge. Predicted travel-time errors through this model were inverted to determine the best-fit 1D velocity structure (see Methods). Seismic ray-paths crossing the caldera were excluded from the inversion, so the resulting model represents average crustal structure away from the volcano. In the final 1D model (not shown), velocity contours 4–6 km s⁻¹ are up to 1 km deeper than on other parts of the Juan de Fuca^{13,14} ridge. The additional 0.5–1.0 km of volcanic extrusives implied by these depressed velocities is not surprising, considering the shallow bathymetry and crustal thickening⁶.

Seismic travel-time prediction was accomplished through a fast ray shooting method performed on a three-dimensional tetrahedral velocity grid. Predicted travel times through the 1D model deviate from real data with an r.m.s. error of 0.181 s. Much of this misfit is contributed by a small number of caldera-crossing rays which are systematically delayed by up to 0.5 s (Fig. 2). A three-dimensional

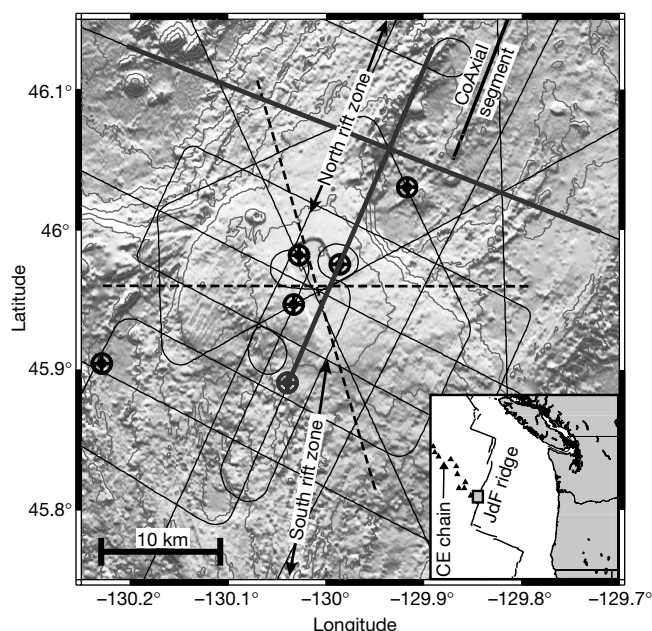


Figure 1 Bathymetry of Axial volcano. Summit caldera, in centre of figure, is at ~1,460 m depth, surrounding sea floor is at ~2,800 m depth, contour interval is 200 m. Thin black lines mark airgun shot lines, crossed circles with central dot show ocean-bottom seismometers used in analysis. Two dashed lines mark cross-sections in Fig. 4. The two crossing thick black lines, one of which ends in a seismometer, display source–receiver geometry for Fig. 2. Scale bar, 10 km. Inset, location of experiment on the Juan de Fuca (JdF) ridge, and the seamounts in the Cobb–Eickelberg (CE) chain.

(3D) tomographic inversion of the travel-time data, starting with the 1D model, was performed to determine a more realistic crustal structure. The final structure derived from the full set of Pg travel times reduced the travel-time misfit to 0.089 s. Checkerboard resolution tests demonstrate that this method and ray geometry are sufficient to resolve the features discussed here (Fig. 3).

A prominent low-velocity zone is centred directly beneath the caldera. It is oval in map view (8 × 12 km elongated northwest–southeast), and has a compressional velocity that is more than 2 km s^{−1} (34%) slower than its surroundings (Fig. 4). The strongest anomaly (>15%) is confined to depths of 2.25–3.5 km below the sea floor. Perturbations of 0.5–1.0 km s^{−1} extend to 5–6 km depth, beyond which the structure is not constrained. The low-velocity feature does not extend under the imaged portion of the rift zones or towards the nearby Vance or CoAxial segments of the Juan de Fuca ridge. The low-velocity anomaly is comparable in depth, twice the width and much larger in amplitude than a similar feature under the Kilauea caldera in Hawaii¹⁵.

Sub-solidus temperature differences can explain some of the low-velocity anomaly under the caldera. We determine the largest possible effect that temperature could have, using an approach similar to that of refs 16 and 17. We assume that the crust beneath the caldera is heated to the 1,150 °C basalt solidus¹⁸ (the highest temperature without the presence of melt); we also assume that the surrounding area has the cooler thermal structure given in Fig. 7f of

Table 1 Parameters used in melt estimates

Parameter	Value	Description
$\partial \ln v_p / \partial T$ at $T < 800$ °C*	$-8.1 \times 10^{-5} \% \text{ K}^{-1}$	Anharmonic effect only ²⁰
$\partial \ln v_p / \partial T$ at $T > 800$ °C	$-16 \times 10^{-5} \% \text{ K}^{-1}$	Anharmonic+anelastic effects ²¹ at 1,150 °C solidus ¹⁸ , $Q = 100$ (ref. 22)
$\partial \ln v_p / \partial F$ (max. melt)	−1.0%	Unrelaxed tubule-shaped inclusions ²⁷
$\partial \ln v_p / \partial F$ (min. melt)	−4.8%	Relaxed random cusped films above $F = 1\%$ (ref. 27)

v_p , P-wave velocity; T , temperature; F , melt fraction.

* Anelastic effects are considered for $T > 800$ °C. Above this temperature Q_{gabbro} drops below 500 (ref. 23), where $1/Q$ is the seismic attenuation. Used in Karato²² equation (6).

ref. 19, for 10 km off the axis of an intermediate-rate spreading ridge. The temperature difference as a function of depth is converted to a velocity perturbation using derivatives of compressional-wave velocity with respect to temperature²⁰ (Table 1). The effect of anelasticity is incorporated for temperatures > 800 °C (refs 21, 22). Up to 0.7 km s^{−1} of the low-velocity anomaly can be explained by thermal effects below the solidus. Errors associated with the thermal assumptions are minimal. Shifting the solidus or the temperature profile by 100 °C changes the velocity effect by < 0.1 km s^{−1}. Microearthquakes in the caldera region, which constrain the depth of the brittle–plastic transition in crustal rheology^{23,24}, do not extend deeper than 2 km, indicating that the magma chamber is overlain by a hot, conductive lid approximately

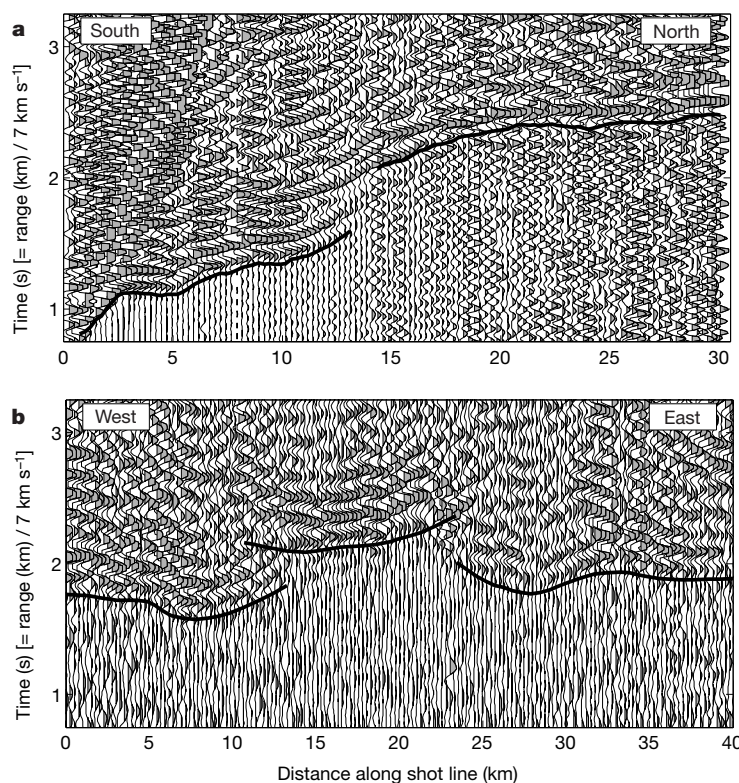


Figure 2 Example record sections showing the low-velocity zone. Shot–receiver geometries are marked on Fig. 1. A bandpass filter and a static topography correction have been applied to each record for display purposes. **a**, Straight line receiver gather. The first-arrival Pg phase is abruptly terminated at a range of 13 km from the receiver. At greater distances, the first significant arrival is delayed by 0.5 s. This phase corresponds to energy refracting through the deeper parts of the low-velocity zone. A similar shadow zone is observed on numerous shot–receiver geometries which cross the caldera. Rays with ranges > 13 km pass unaffected above the low-velocity zone. This record demonstrates that the low-velocity zone begins at depth beneath the caldera, and does not extend into the shallowest crust. **b**, Fan shot receiver gather. Shot–receiver ranges

are 20–29 km (too close for P_mP Moho reflection in the region). Times have been adjusted for range. The effect of the low-velocity anomaly is evident in the centre of the record. First-arrival Pg phases are clear for rays which pass to either side of the caldera region: shots at 0–13 km and 23–40 km on this plot. Between 13 and 26 km, this phase becomes delayed and quickly falls below the noise level. A strong second arrival is visible with a delay of ~0.5 s. This large-amplitude phase is the same one found on record **a** at ranges > 13 km. This record reveals the lateral extent of the low-velocity zone. Thick black lines indicate modelled traveltimes. Time axes are reduced by 7 km s^{−1} for display purposes.

0.5–1.0 km thick. Thus the crust is probably hotter than our model, and our treatment of the effect of temperature on seismic velocity is conservative (that is, an overestimate).

After removing thermal effects from the velocity-anomaly model, the remaining low velocity is attributed to melt (Fig. 4d). The influence of melt content on velocity depends on the distribution of the melt and assumptions about the principal mechanics. Estimates of velocity as a function of melt fraction have been obtained from numerical^{25–27} and laboratory²⁸ experiments. We use the recent relations of ref. 27, which generally find smaller melt contents than other models. Our minimum estimates of magma content assume that melt occurs in thin cusped films under relaxed conditions (that is, melt ‘squirts’ equalizes pressure between connected inclusions) (Table 1). Maximum melt estimates assume spherical melt inclusions.

Significant concentrations of melt beneath Axial delineate a magma reservoir. The greatest melt concentrations of 5–25%, depending on the model, occur 2.5–3.5 km below the caldera, consistent with the centre of deformation during the 1998 eruption^{3,4}. At this depth the width of the reservoir is 8×12 km,

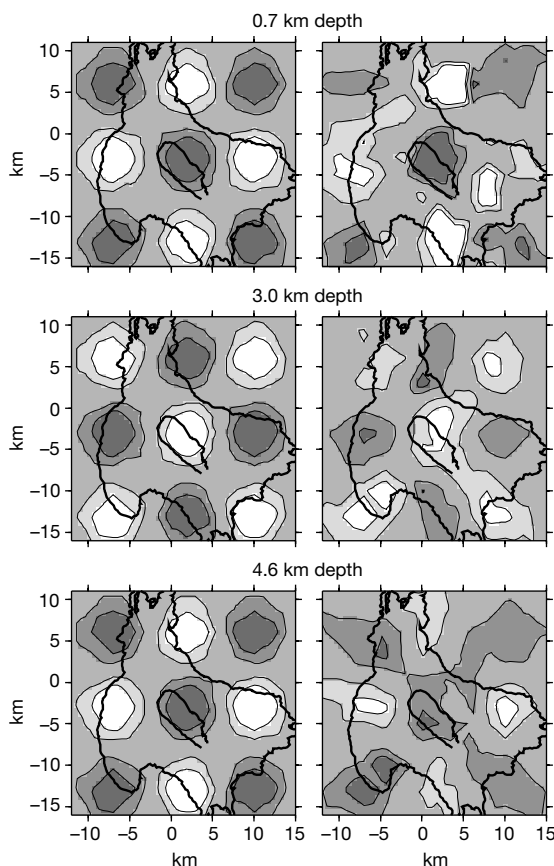


Figure 3 Results of checkerboard resolution tests in a 25×25 km box. The synthetic input model (left panels) was created by adding a grid of alternating negative and positive gaussian-shaped anomalies (5%) to the 1D starting model. Travel times were forward-modelled by ray-tracing through this structure using a geometry identical to the one in the real tomographic inversion and assuming similar errors. The same inversion strategy was then applied to the synthetic data set to see how well the input structure could be recovered (right panels). Three depth sections are shown here. The shallow resolution (0.7 km below sea floor) is noisy, though general features are recovered. At depths of the low-velocity zone (3.0 km) the anomalies are well-located. Peak amplitudes are somewhat less than in the input model as a result of smearing. Deeper in the crust (4.6 km) the overall geometry is recovered though smearing is more evident. In the middle and lower crust under the caldera, this test demonstrates our ability to resolve features similar to the observed magma chamber.

elongated similarly to the caldera. We consider these minimum melt contents, as a thin lens of near-pure melt could go undetected owing to intrinsic diffraction effects²⁹. Small amounts of melt, which extend to the bottom of image resolution at 6 km below the sea floor, contribute significantly to the total melt content. The presence of melt in the deep crust suggests either that replenishment of the magma chamber is continuing, or that the crust has yet to

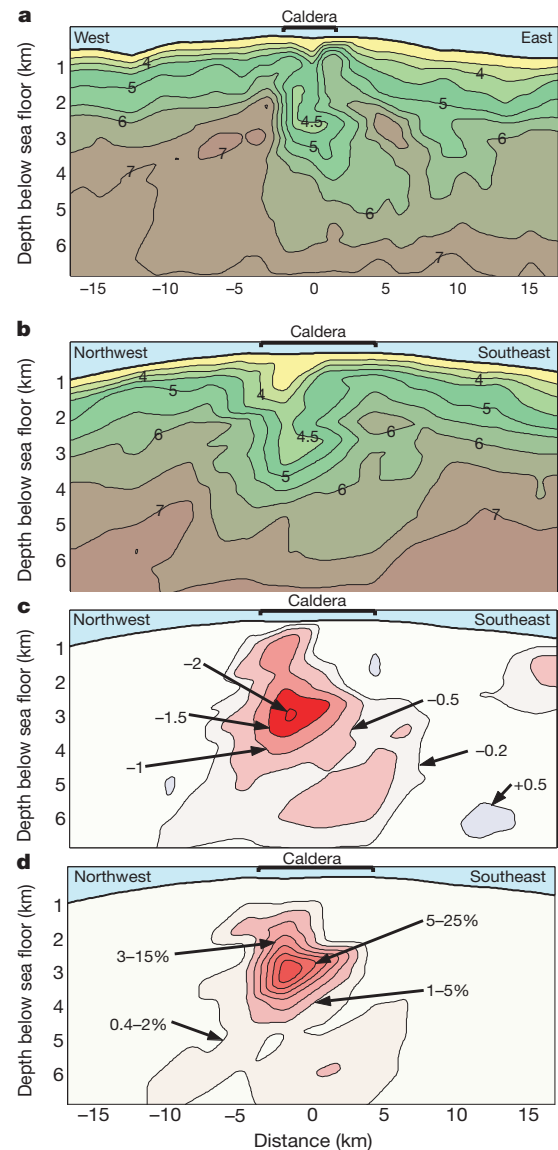


Figure 4 Cross-sectional views of Axial volcano, showing velocity structure and interpretation. **a**, East–west and **b**, northwest–southeast profiles through caldera, as marked in Fig. 1. Compressional-wave velocity is contoured in intervals of 0.5 km s^{-1} . The caldera width is indicated above each panel. Depths are relative to the caldera floor, which is 1.46 km below sea level. The low-velocity zone directly beneath the caldera is the result of a partially molten magma body. These panels show roughly radial symmetry beneath Axial volcano, with features being somewhat elongated along the caldera. The low-velocity zone is not seen to extend away from the caldera in any direction. It does not extend along the rifts or towards the adjacent ridge segments. **c**, Velocity anomaly plot of the northwest–southeast profile. Contours show the difference between the final 3D velocity structure and the 1D bathymetry-draped starting model that characterizes the structure away from the caldera. **d**, After removing the velocity anomaly that could be due to thermal differences, the remaining anomaly is attributed to melt. The different shape of the melt region in **d** and the velocity anomaly in **c** is due to the removal of thermal effects. Comparison with the modelled geothermal gradient removes more of the velocity anomaly at shallower depths.

cool below the solidus following the last replenishment from the mantle.

Even conservative models predict several cubic kilometres of magma beneath Axial volcano. The total melt volume is found by summing the partial melt in the reservoir. The end-member models discussed above yield total magma contents of 5–21 km³ spread throughout a 250-km³ reservoir. Because the bottom of this reservoir is not imaged here, the total volume may be larger still. The amount of magma can best be understood by comparing it to the volume emplaced during eruptions. The magma currently stored beneath Axial volcano is 25–200 times larger than the 0.1–0.2 km³ total estimated volume of the 1998 eruption^{3,4}. There is far more magma than can be removed by several eruptions, implying that stored magma is a long-lived feature at Axial even between eruptive cycles. Individual eruptions do not empty, or even significantly deplete, the reservoir. The small observed melt fractions probably rule out large-scale convective overturn in the magma chamber. However, Axial erupts remarkably homogeneous basalts compared to nearby segments of the Juan de Fuca ridge³⁰, suggesting that some mixing does occur. If mixing occurs in even a portion of the magma reservoir, the smoothing effect on chemical variations could be significant because the melt content is so much larger than the typical eruption size. In light of the lengthy implied residence time, some degree of mixing seems possible.

The residence time in the crust depends on the total magma content and the flux. The observation of large amounts of melt under Axial volcano bolsters previous observations that the volcano provides much of the magma required to accommodate spreading across its rift zones to the north and south. The dimensions of the rift zones can be used to estimate the flux of magma through the Axial system. Assuming that Axial's 100-km rift system is fed primarily from this central magma chamber, and that the rifts are opening at 5–6 cm yr⁻¹ as is the rest of the ridge, an event similar to the one in 1998 is needed on the order of every ten years. If either of these assumptions is incorrect, the flux will decrease accordingly. Combining the flux with the reservoir estimates suggests an average storage time for magma in the crust of a few hundred to a few thousand years, and possibly much longer. The long residence time for melt beneath the discrete volcanic centre at Axial volcano demonstrates that eruptions are a small perturbation to a much larger semi-permanent magma storage system, which is unconnected to nearby segments of the Juan de Fuca ridge. □

Methods

Tomographic inversion

The velocity field is parametrized on an irregular tetrahedral grid, with horizontal node spacing of 2 × 2 km near the caldera and greater spacing in poorly constrained areas of the model. The grid is vertically sheared to conform to the bathymetry. Average vertical node spacing is 0.4 km in the upper crust, increasing to 1.0 km in the lower part of the model. The water layer and bathymetry, as derived from multibeam sonar, are explicitly prescribed in the model. No water path 'corrections' are required. Velocity within each tetrahedron is a linear gradient of the values specified at the four vertices. Travel times and their Fréchet derivatives as a function of node velocities are calculated by summing the analytic ray-path solutions in each tetrahedron. A linearized damped least-squares inversion, parametrized on the same grid, is used to update the velocity field. As the inversion is parametrized in velocity (as opposed to slowness), which increases from < 4 km s⁻¹ at the top of the model to > 7 km s⁻¹ at the bottom, the damped least-squares approach preferentially accommodates model variation at shallow depths. This counters the tendency of shallow and mid-crustal features to be smeared into deeper crust.

The bathymetrically draped 1D model was improved by applying a 1D constraint to the inversion. Travel times for rays that cross the caldera region were excluded, so the velocity structure represents average crust away from the active volcanic edifice. This was the starting model for 3D inversions and the reference model against which anomalies are assessed.

The shadow zones illustrated in Fig. 2 provide much of the 0.181-s r.m.s. error in the 1D travel-time predictions. Three iterations of the inversion using the full set of Pg travel times reduced this misfit to 0.089 s. An empirically derived inversion damping of 0.11|G^TG|_∞, where G is the matrix of Fréchet derivatives, G^T is its matrix transpose, and ||...||_∞ is the L_∞ matrix norm, achieved the greatest misfit reduction while introducing the least model complexity (as measured by the variance of the difference between the final and starting models). Although further reductions could be achieved with decreased inversion

damping and more iterations, the small improvements did not warrant the greater model complexity.

The model resolution was assessed by inverting travel times from synthetic input models using an identical ray geometry and the same inversion strategy as above. Synthetic models contained a grid of 3D gaussian-shaped anomalies of alternating sign. These tests demonstrate our ability to resolve features 5–10 km in diameter to depths of 5–6 km below the sea floor in the vicinity of the caldera (Fig. 3).

Received 30 July; accepted 10 September 2001.

- Johnson, H. P. & Embley, R. W. Axial seamount: An active ridge axis volcano on the central Juan de Fuca ridge. *J. Geophys. Res.* **95**, 12689–12696 (1990).
- Embley, R. W., Chadwick, W. W. Jr, Clague, D. A. & Stakes, D. 1998 eruption of Axial Volcano; multibeam anomalies and seafloor observations. *Geophys. Res. Lett.* **26**, 3425–3428 (1999).
- Chadwick, W. W. Jr, Embley, R. W., Milburn, H. B., Meinig, C. & Stapp, M. Evidence for deformation associated with the 1998 eruption of Axial Volcano, Juan de Fuca ridge, from acoustic extensometer measurements. *Geophys. Res. Lett.* **26**, 3441–3444 (1999).
- Fox, C. G., Chadwick, W. W. Jr & Embley, R. W. Direct observation of a submarine volcanic eruption from a sea-floor instrument caught in a lava flow. *Nature* **412**, 727–729 (2001).
- Desonie, D. L. & Duncan, R. A. The Cobb-Eickelberg seamount chain; hotspot volcanism with mid-ocean ridge basalt affinity. *J. Geophys. Res.* **95**, 12697–12711 (1990).
- Hooff, E. E. & Detrick, R. S. Relationship between axial morphology, crustal thickness, and mantle temperature along the Juan de Fuca and Gorda Ridges. *J. Geophys. Res.* **100**, 22499–22508 (1995).
- Tolstoy, M., Vernon, F. L., Orcutt, J. A. & Wyatt, F. K. The breathing of the seafloor: Tidal correlations of seismicity at Axial Volcano. *Geology* (submitted).
- Dziak, R. P. & Fox, C. G. Long-term seismicity and ground deformation at Axial Volcano, Juan de Fuca Ridge. *Geophys. Res. Lett.* **26**, 3641–3644 (1999).
- Dziak, R. P. & Fox, C. G. The January 1998 earthquake swarm at Axial Volcano, Juan de Fuca Ridge: Hydroacoustic evidence of seafloor volcanic activity. *Geophys. Res. Lett.* **26**, 3429–3432 (1999).
- Fox, C. G. In situ ground deformation measurements from the summit of Axial Volcano during the 1998 volcanic episode. *Geophys. Res. Lett.* **26**, 3437–3440 (1999).
- Hildebrand, J. A. et al. A seafloor and sea surface gravity survey of Axial Volcano. *J. Geophys. Res.* **95**, 12751–12763 (1990).
- Webb, S. C., Deaton, T. K. & Lemire, J. C. A broadband ocean-bottom seismometer system based on a 1-Hz natural period geophone. *Bull. Seismol. Soc. Am.* **91**, 304–312 (2001).
- Cudrak, C. F. & Clowes, R. M. Crustal structure of Endeavour Ridge segment, Juan de Fuca Ridge, from a detailed seismic refraction survey. *J. Geophys. Res.* **98**, 6329–6349 (1993).
- Sohn, R. A., Webb, S. C., Hildebrand, J. A. & Cornuelle, B. D. Three-dimensional tomographic velocity structure of upper crust, CoAxial segment, Juan de Fuca Ridge; implications for on-axis evolution and hydrothermal circulation. *J. Geophys. Res.* **102**, 17679–17695 (1997).
- Dawson, P. B., Chouet, B. A., Okubo, P. G., Villaseñor, A. & Benz, H. M. Three-dimensional velocity structure of the Kilauea caldera. *Geophys. Res. Lett.* **26**, 2805–2808 (1999).
- Dunn, R. A., Toomey, D. R. & Solomon, S. C. Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the East Pacific Rise at 9°30' N. *J. Geophys. Res.* **105**, 23537–23555 (2000).
- Canales, J. P., Collins, J. A., Escartin, J. & Detrick, R. S. Seismic structure across the rift valley of the Mid Atlantic Ridge at 23°20' (MARK area): Implications for crustal accretion processes at slow spreading ridges. *J. Geophys. Res.* **105**, 28411–28425 (2000).
- Sinton, J. M. & Detrick, R. S. Mid-ocean ridge magma chambers. *J. Geophys. Res.* **97**, 197–216 (1992).
- Henstock, T. J., Woods, A. W. & White, R. S. The accretion of oceanic crust by episodic sill intrusion. *J. Geophys. Res.* **98**, 4143–4161 (1993).
- Christensen, N. I. Compressional wave velocities in rocks at high temperatures and pressures, critical thermal gradients, and crustal low-velocity zones. *J. Geophys. Res.* **84**, 6849–6857 (1979).
- Karato, S. Importance of anelasticity in the interpretation of seismic tomography. *Geophys. Res. Lett.* **20**, 1623–1626 (1993).
- Kampfmann, W. & Berckhemer, H. High temperature experiments on the elastic and anelastic behaviour of magmatic rocks. *Phys. Earth Planet. Inter.* **40**, 223–247 (1985).
- Chen, W. P. & Molnar, P. Focal depths for intracontinental and intraplate earthquakes and their implications for the thermal and mechanical properties of the lithosphere. *J. Geophys. Res.* **88**, 4183–4214 (1983).
- Huang, P. Y. & Solomon, S. C. Centroid depths of mid-ocean ridge earthquakes: dependence on spreading rate. *J. Geophys. Res.* **93**, 13445–13477 (1988).
- Schmeling, H. Numerical models on the influence of partial melt on elastic, anelastic and electric properties of rocks. Part I: elasticity and anelasticity. *Phys. Earth Planet. Inter.* **41**, 34–57 (1985).
- Mainprice, D. Modelling the anisotropic seismic properties of partially molten rocks found at mid-ocean ridges. *Tectonophysics* **279**, 161–179 (1997).
- Hammond, W. C. & Humphreys, E. D. Upper mantle seismic wave velocity: Effects of realistic partial melt geometries. *J. Geophys. Res.* **105**, 10975–10986 (2000).
- Sato, H., Sacks, I. S. & Murase, T. The use of laboratory velocity data for estimating temperature and partial melt fraction in the low-velocity zone; comparison with heat flow and electrical conductivity studies. *J. Geophys. Res.* **94**, 5689–5704 (1989).
- Wilcock, W. S. D., Dougherty, M. E., Solomon, S. C., Purdy, G. M. & Toomey, D. R. Seismic propagation across the East Pacific Rise: finite difference experiments and implications for seismic tomography. *J. Geophys. Res.* **98**, 19913–19932 (1993).
- Rhodes, J. M., Morgan, C. & Liias, R. A. Geochemistry of Axial Seamount lavas; magmatic relationship between the Cobb hotspot and the Juan de Fuca Ridge. *J. Geophys. Res.* **95**, 12713–12733 (1990).

Acknowledgements

We thank C. Golden, J. Floyd, D. Bohnenstiehl, V. Ballu, and the crews of the RV *Thomas G. Thompson* and the RV *Maurice Ewing* for efforts at sea; M. Spiegelman, J. Chadwick, M. Perfit and S. Carbotte for discussions; and C. Fox and W. Chadwick for comments on the manuscript. This work was supported by the US National Science Foundation.

Correspondence and requests for materials should be addressed to M.W. (e-mail: mwest@ldeo.columbia.edu).

Organic chemistry of embalming agents in Pharaonic and Graeco-Roman mummies

Stephen A. Buckley & Richard P. Evershed

Organic Geochemistry Unit, Biogeochemistry Research Centre, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

Chemical treatments were an essential element of ancient Egyptian mummification. Although the inorganic salt natron is recognized as having a central role as a desiccant¹, without the application of organic preservatives the bodies would have decomposed in the humid environment of the tombs². The nature of the organic treatments remains obscure, because the ancient Egyptians left no written record of the process. Secondary textual evidence for mummification is provided by Herodotus³, Diodorus Siculus⁴, Strabo⁵ and Pliny⁶. The most important

account is that of Herodotus³ (about 450 yr BC), although archaeological evidence shows that by this time the process had declined significantly and the best results had been achieved centuries before⁷. His account mentions myrrh, cassia, palm wine, 'cedar oil' (still widely disputed^{8–10}) and 'gum'; however, it is vague with respect to the specific natural products used. Here we report the results of chemical investigations of a substantial collection of samples of tissues, wrappings and 'resinous/bituminous' materials from provenanced and dated Egyptian mummies. We focused on examples of the 'classic' mummy-making culture of the Pharaonic or dynastic period, from which we can begin to track the development of mummification chronologically.

Studies of the organic materials used in mummification were carried out earlier in the last century (for example, see refs 11–13), but the analytical techniques available limited their ability to identify aged, complex organic mixtures (that is, the 'resins', spices, and so on). To gain a proper understanding of the mummification process and its development, it is vital to study securely provenanced and dated mummies using modern investigative chemical techniques. Only four such studies (which includes one performed in our own laboratory) have been carried out in recent

Table 1 Provenance and date of mummies, origin of balm samples and their chemical composition

Mummy†	Date/age	Provenance	Sample location and description#	Inferred components of embalming 'resin' ☆	Relative abundance (%)**
Male adult* ('Khnumnakht') 21471	1985–1795 yr BC (XII dynasty)	Rifeh	Resin/tissue/bandaging	Fat/oil ^g Proteinaceous material	90% 10%
Female adult† 1909.527	1650–1550 yr BC (XVII dynasty)	Qurna, Thebes	Resin/tissue from head of right tibia	Fat/oil ^g Coniferous pitch Balsam (?)	99% 1% Trace
Child (sex ?)† 1909.527	1650–1550 yr BC (XVII dynasty)	Qurna, Thebes	Unspecified bone and cartilage	Fat/oil ^g Coniferous pitch Balsam (?)	82% 1% 16%
Head (sex ?)‡ 1976.159.267	1550–1069 yr BC (XVIII–XX dynasties)	Thebes	Skin/resin from back of cranium	Fat/oil ^g A sugar/gum Coniferous resin Balsam (?)	95% 0.3% 2% 2%
Male adult§ ('Horemkenesi') Ha 7386	1069–945 yr BC (XXI dynasty)	Deir el Bahri, Thebes	Resinous material from left side of top of spine	Fat/oil ^g	100%
Female(?) adult EA74303	1069–664 yr BC (XXI–XXV dynasties)	Thebes?	Resin from chest cavity	Fat/oil ^k Coniferous resin/pitch Balsam (?) Beeswax ^{n,s}	61% 6% 1.5% 31%
Female adult§ ('Neskhons') H 5062	945–715 yr BC (XXII dynasty)	Thebes	Resin-soaked outer wrapping from left side of neck	Plant oil ^d Coniferous resin Triterpenoid resin ? Balsam (?) Wax	82% 0.5% 4% 0.9% 11%
Male adult‡ ('Pedeamun') 1953.72	664–404 yr BC (XXVI–XXVII dynasties)	Thebes	Resin from top of cranium	Fat/oil ^m Coniferous resin Balsam (?) Beeswax ^{o,t}	61% 0.2% 1.5% 38%
Female adult† 1956.352	332–30 yr BC (Ptolemaic)	Thebes	Resin attached to linen thread on right ankle	Fat/oil ^f Pistacia resin Balsam (?) Beeswax ^{u,v}	6% 6% 0.4% 87%
Male adult (1)† 1911.2101	30 yr BC to AD 395 (Roman)	Hawara	Resin-soaked outer wrapping below right scapula	Fat/oil ^f Coniferous resin Balsam (?) Beeswax ^{u,v}	78% 16% Trace 6%
Male adult (2)† 1911.2102	30 yr BC to AD 395 (Roman)	Hawara	Resin from side/base of left foot	Fat/oil ^g Coniferous resin Balsam (?) Beeswax ^w	35% 37% 0.2% 27%
Male child (wrapped)† 1956.357b	30 yr BC to AD 395 (Roman)	Thebes	Resin on outer wrapping from area between calves	Fat/oil ^h Coniferous resin	90% 9%
Male child (unwrapped)† 1956.357c	30 yr BC to AD 395 (Roman)	Thebes	Resin from abdominal cavity (kidney area)	Fat/oil ^f Coniferous resin Balsam (?)	87% 13% 0.1%

Several samples were taken from all but one mummy (unwrapped male child). For example, 15 different samples were analysed from various parts of the body of Horemkenesi.

* Manchester Museum; † National Museum of Scotland; ‡ Liverpool Museum; § Bristol Museum; || British Museum.

† Museum number.

The term 'resin' denotes physical appearance and does not presuppose any chemical composition or biological origin.

☆ Superscript letters (a–w) refer to the histograms shown in Fig. 1.

** Percentage relative abundance based on absolute concentrations, calculated on the basis of internal standards added at the extraction stage. Compositions do not imply that they were the original formulations of the embalmers, owing to possible chemical changes over time.

times^{14–17}.

This study involves the first systematic chemical investigation of a collection of provenanced Egyptian mummies dating from the mid-dynastic period (c. 1,900 yr BC) to the late Roman period (AD 395), which encompasses the period at which the mummification process was at its peak (1,350–1,000 yr BC). We have used diagnostic marker compounds that are present in the original 'balms', are resistant to degradation and can be related to specific embalming agents. Gas chromatography with mass spectrometry (GC-MS) and thermal desorption (TD)- or pyrolysis (Py)-GC-MS facilitate the molecular separation and identification of the marker compounds.

Because of the probable nature, possible processing and burial history of embalming materials, both free and polymerized components are likely to be present. Therefore, this 'dual' approach allows the characterization and identification of both the free (solvent-extractable) marker compounds and the recognizable subunits of polymeric materials not amenable to the more conventional GC-MS approach¹⁷. Given the understandable sensitivity surrounding the study of mummies, the amount of sample required is an important consideration. Sequential TD-GC-MS and Py-GC-MS require very small sample sizes (<0.1 mg), facilitating the virtually non-destructive analysis of valuable museum specimens.

A summary of the results of this study is given in Table 1. The main products seen in all mummies are degraded acyl lipids, which are derived from plant oils and animal fats (Fig. 1a–m). Where

samples were taken directly from the bodies, these may well derive from endogenous body lipids, such as triacylglycerols from fats or cell-membrane phospholipids. But even after accounting for decay, for the most part these display non-human fatty-acid distributions. In cases where degraded fats or oils are, for example, seen in the bandaging and are not directly in contact with the body, they must derive from their intentional application as part of the embalming process.

In most cases, the fatty-acid distributions (Fig. 1) are indicative of plant origins; that is, they have a high abundance of C_{16:0} compared with their abundance of C_{18:0}. But it is not possible to assign a precise plant origin on the basis of fatty acids alone, owing to degradation of the major unsaturated components; such degradation is evident from the presence of mono- and dihydroxycarboxylic acids, which are indicative of autoxidation. The presence of plant oils (and to a lesser extent animal fats) suggests that they were key ingredients in mummification, and were probably used as a less-costly base with which to mix and apply more exotic embalming agents to the bodies and/or wrappings.

Their widespread use indicates that the embalmers were aware of the special properties of unsaturated oils and fats that allow them to 'dry', or rather, to polymerize spontaneously¹⁸ (later also appreciated by the oil painters of western Europe). This polymerization would have produced a highly crosslinked aliphatic network, which would have stabilized otherwise fragile tissues and/or textile

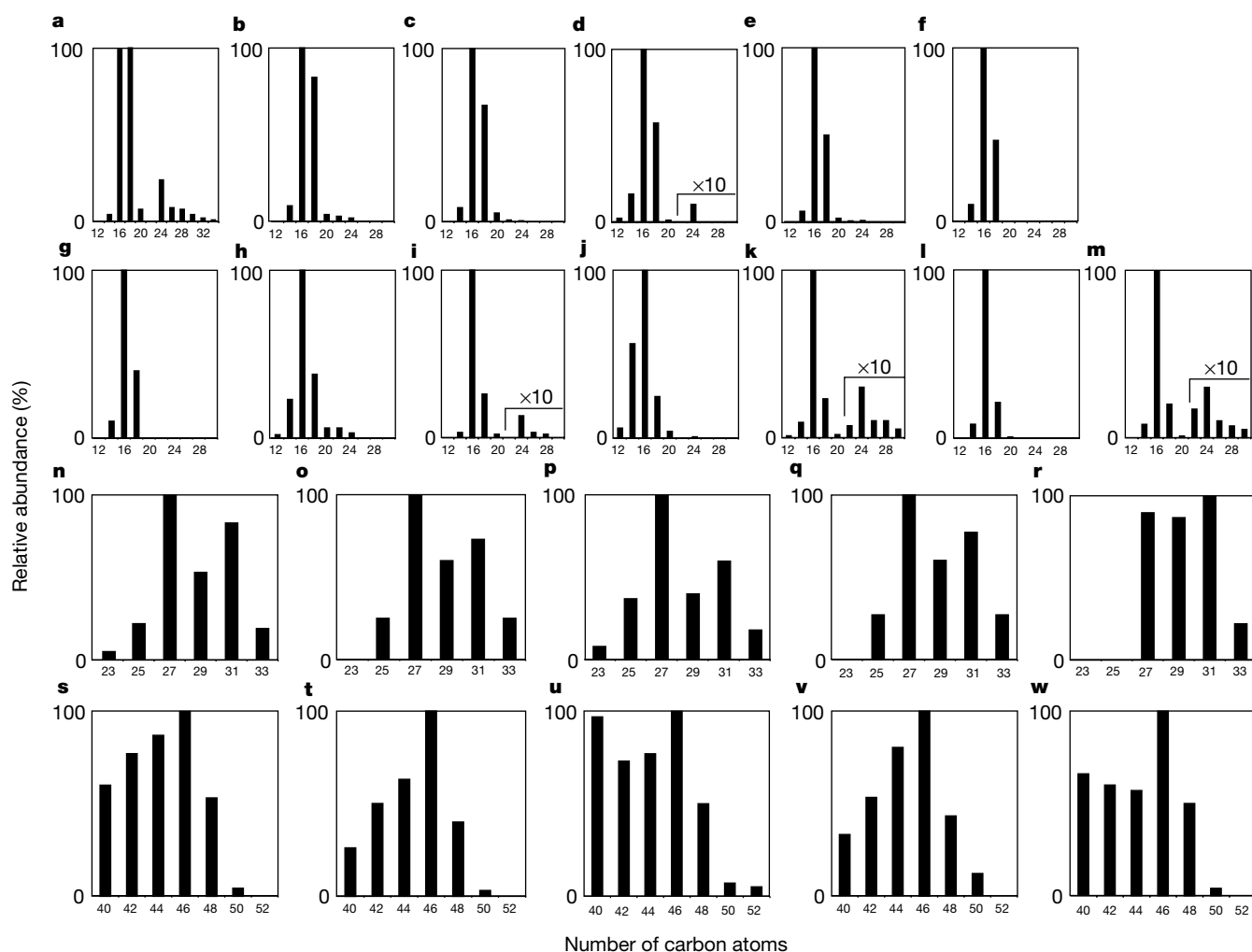


Figure 1 Chemical analyses of samples taken from the mummies in this study. Histograms show the distribution of fatty acids of equal carbon number in acid fractions (a–m; see Table 1), hydrocarbons of odd carbon number in neutral fractions from

samples containing beeswax (n–r), wax esters of even carbon number in neutral fractions from samples containing beeswax (s–w).

wrappings against degradation by producing a physico-chemical barrier that impedes the activities of microorganisms.

The results from the wide date range of the mummies investigated reveal notable trends in the evolution of the use of other commodities (Fig. 2). For example, beeswax and coniferous resin clearly increase in their prominence through time, and are found in material taken both directly from the bodies and from the wrap-

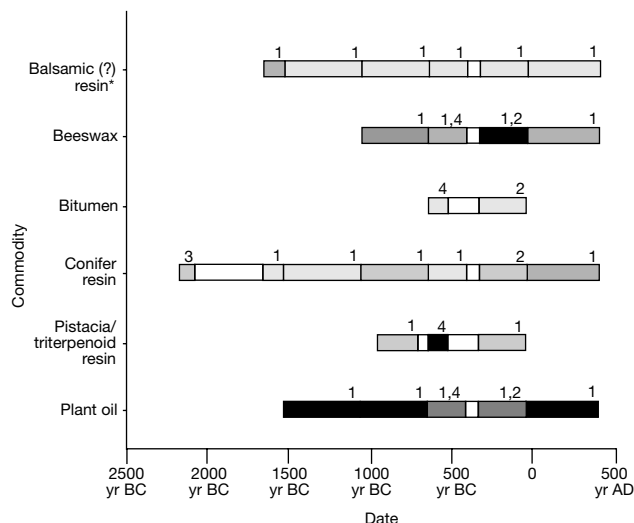


Figure 2 The use of organic commodities (preservatives/unguents) identified in chemical investigations of the 'balms' of provenanced and dated ancient Egyptian mummies. Grey scale reflects the relative abundance of a particular commodity found in the balm samples, with darker shades representing a higher relative abundance. Asterisk, tentative identification of balsamic resins; 1, data from this study; 2, data from ref. 14; 3, data from ref. 16; 4, data from ref. 15.

pings. Coniferous resin is identified by the presence of both functionalized and defunctionalized diterpenoid components. For example, 7-oxodehydroabietic acid and 15-hydroxy-7-oxodehydroabietic acid were usually the dominant diterpenoid components, and the normally abundant dehydroabietic acid was virtually absent.

Although coniferous resins were clearly used in the embalming process at least as early as 2,200 yr BC (VI dynasty)¹⁶, their use becomes most apparent in later periods; both the tissues and the wrappings of mummies from the Roman period (30 yr BC to AD 395) contain appreciable quantities (up to 37%) of coniferous diterpenoids. The increasing use of coniferous resin suggests that the embalmers may have become aware of the ability of specific natural products to inhibit microbial degradation by means of mechanisms (physico-chemical barriers and antimicrobial action) analogous to their protective roles in the plants from which they derived.

A significant trend is also seen in the use of beeswax, which is characterized chemically by alkanes (C_{25} – C_{33}) (Fig. 1n–r), wax esters (C_{40} – C_{50}) (Fig. 1s–w) and hydroxy wax esters (C_{42} – C_{54}). It first appears notably later than coniferous resin, with its positive identification in a resinous coating taken from the chest cavity of a female mummy of the Third Intermediate Period (XXI to XXV dynasty; 1,069–664 yr BC). In a sample taken from 'Pedeamun', a XXVI dynasty (664–525 yr BC) mummy, the solvent-soluble extracts comprise 38% beeswax (Fig. 3). But an even greater amount was found in a very extensive black 'resinous' layer from a female mummy of Ptolemaic date (332–30 yr BC) (Fig. 4), in which beeswax made up 87% of the solvent extractable components; in other words, this treatment was essentially beeswax mixed with a small amount of *Pistacia* resin (see also below). The choice of beeswax was clearly motivated by its hydrophobic/sealant and antibacterial properties, notwithstanding its symbolic significance¹⁹.

Other components of the embalming 'resin' are triterpenoids and hydroxyaromatic acids, which are diagnostic of plant products. The

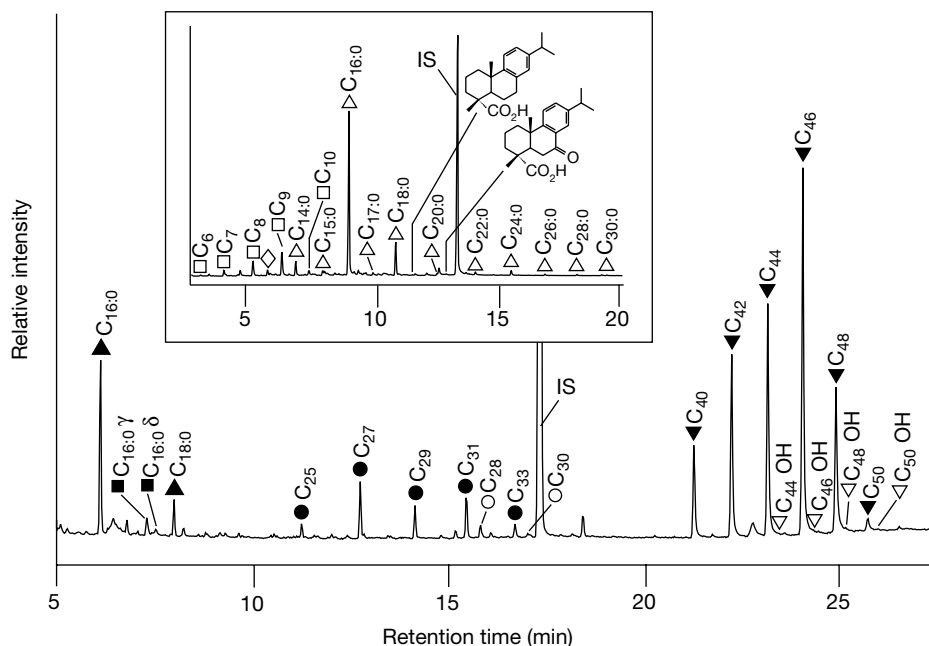


Figure 3 Reconstructed total-ion chromatogram of the trimethylsilylated neutral fraction (after solvent extraction and fractionation) of 'resin' from the top of the cranium of the Late period male adult 'Pedeamun', XXVI to XXVII dynasty (664–404 yr BC). Peak identities (x indicates carbon chain length): filled triangles, $C_{x:0}$ indicates saturated fatty acid methyl esters; filled squares, $C_{x:0}$ indicates saturated lactones (carbon chain length x is followed by the position of the lactone ring); filled circles, C_x indicates alkanes; open circles, C_x indicates alkanols; filled inverted triangles, C_x indicates wax esters; open inverted

triangles, C_x OH indicates hydroxy wax esters; IS, internal standard (*n*-tetratriacontane). Inset displays a reconstructed total-ion chromatogram of the trimethylsilylated acid fraction (after solvent extraction and fractionation) of this sample. Peak identities: open triangles, $C_{x:0}$ indicates saturated fatty acids; open squares, C_x indicates α,ω -dicarboxylic acids; open diamond, vanillic acid. Also shown are the structures of two diterpenoid acids identified: dehydroabietic acid and 7-oxodehydroabietic acid. IS, internal standard (*n*-heneicosanoic acid).

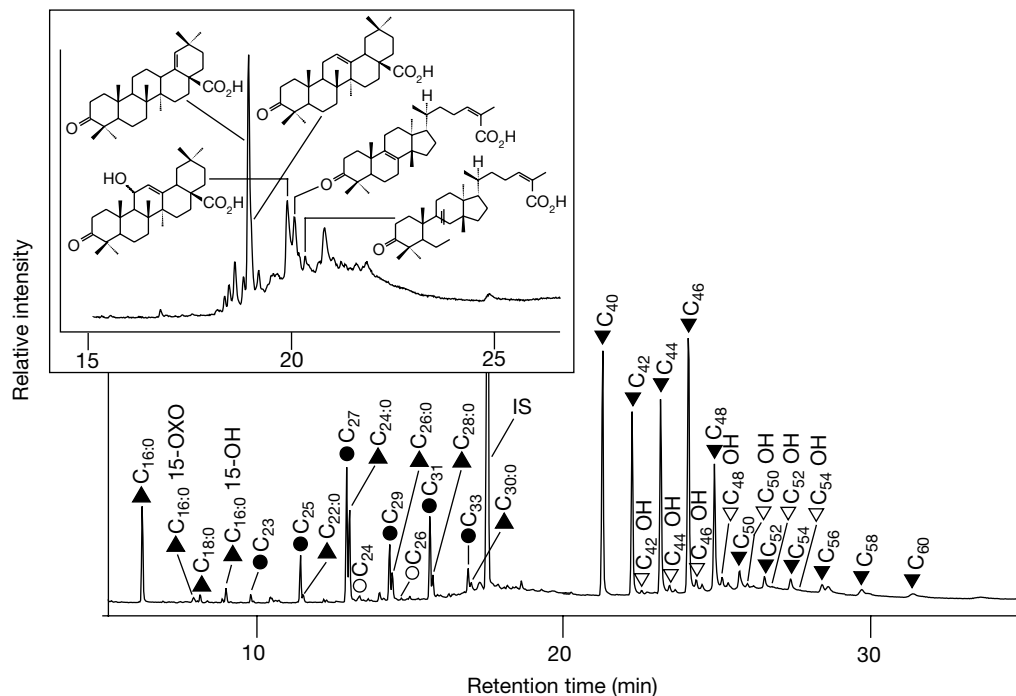


Figure 4 Reconstructed total-ion chromatogram of the trimethylsilylated neutral fraction (after solvent extraction and fractionation) of 'resin' attached to linen thread on the right ankle of a female adult from the Ptolemaic period (332–30 yr BC). Peak identities (x indicates carbon chain length): filled triangles, C_{x0} indicates saturated fatty acid methyl esters; filled circles, C_x indicates alkanes; open circles, C_x indicates alkanols; filled inverted triangles, C_x indicates wax esters; open inverted triangles, C_x OH indicates

hydroxy wax esters; IS, internal standard (*n*-tetratriacontane). Inset displays a reconstructed total-ion chromatogram of the trimethylsilylated acid fraction (after solvent extraction and fractionation) of this sample, showing the structures of the *Pistacia* triterpenoids identified: moronic acid, oleanonic acid, hydroxyoleanonic acid, isomasticadienonic acid and masticadienonic acid.

former include the isomasticadienonic, masticadienonic, moronic and oleanonic acids that are diagnostic of the presence of *Pistacia* resin¹⁸ and are found in the Ptolemaic female mummy (Fig. 4). The *Pistacia* content is significant, constituting 6% of the extractable lipids. Triterpenoid acids, including a prevalence of dammaranes that can occur in highly degraded *Pistacia* resins²⁰, were present in significant quantities (4%) in the wrappings of the XXII dynasty (945–715 yr BC) mummy 'Neskhnos'. However, the absence of moronic acid (which is used often to characterize *Pistacia*) leaves its precise botanical origin uncertain.

It is less easy to assign an origin to the hydroxy aromatic acids, but these are clearly plant derived. Although sawdust was often used to pack the cavities of mummies, the derivation of these acids from lignin is unlikely as there were no detectable lignin building-block compounds (methyl, ethyl, *n*-propyl and vinyl guaiacols) in any of the Py-GC-MS analyses. Present in ten of the mummies, their distribution is generally dominated by 4-hydroxy-3-methoxybenzoic acid with smaller amounts of *p*-hydroxybenzoic acid, which is consistent with oxidation of either unsaturated C-3 side chains (either carboxyl or hydroxyl substituted) in the aromatic benzoate ester of balsamic resins or ferulic acid—a principal component of *Umbelliferae*²¹.

The absence of the precursor compounds is not unexpected, given the susceptibility of these to the oxidation that is prevalent in mummies. Whereas low abundances of the components may originate from dissolved organic matter in the Nile water used in washing the bodies, higher concentrations, such as 16% in the XVII dynasty child mummy, would be consistent with a plant-resin treatment. The choice of these latter natural products by the ancient embalmers again points to their recognition of the antibacterial properties²² of these substances and hence their vital role in the mummification process.

Our results show that a mixture of commodities, with a composi-

tional diversity greater than reported previously, was used. The use of drying oils was clearly widespread, with coniferous resins and beeswax increasing in importance with time. Notably, despite the frequent and rather indiscriminate use of the term 'bitumen' in connection with mummification, the idea that its general use in embalming has now been proved "unequivocally"²³ cannot be supported. In repeated searches, the steranes and hopanes characteristic of petroleum bitumens^{14,24} were not detected, which leads us to the conclusion that natural bitumens were not used in these mummies.

Furthermore, given that beeswax generally constitutes the greater proportion of the organic materials rather than the resins, the term embalming 'wax' has at least equal validity to so-called 'resin'. In this respect it is especially interesting to note that the Coptic (directly derived from the ancient Egyptian language) word for wax is in fact 'mum'²⁵. These results clearly show that, despite statements to the contrary^{10,23}, we do not know all there is to know about Egyptian mummification, and caution needs to be exercised when making assumptions about the materials that the ancient Egyptian embalmers might have used. □

Received 30 May; accepted 31 August 2001.

- Sandison, A. T. The use of natron in mummification in ancient Egypt. *J. Near Eastern Studies* **22**, 259–267 (1963).
- Hunter, J., Roberts, C. & Martin, A. in *Studies in Crime: An Introduction to Forensic Archaeology* 63–78 (Batsford, London, 1996).
- Herodotus *The Histories* (trans. De Séincourt, A.) Book II, 160–161 (Penguin, Harmondsworth, 1954).
- Diodorus Siculus *The Library of History* (trans. Oldfather, C. H.) (Heinemann, London, 1935).
- Strabo *The Geography* (trans. Jones, H. L.) (Heinemann, London, 1932).
- Pliny (the Elder) *Natural History* (trans. Rackham, H.) (Heinemann, London, 1968).
- Smith, G. E. & Dawson, W. R. *Egyptian Mummies* (Allen & Unwin, London, 1924).
- Baumann, B. B. The botanical aspects of ancient Egyptian embalming. *Econ. Bot.* **14**, 84–104 (1960).
- Lucas, A. "Cedar"—tree products employed in mummification. *J. Egypt. Archaeol.* **XVII**, 13–21 (1931).
- Mendelsohn, S. The mortuary crafts of ancient Egypt. *CIBA Symposia* 1795–1804 (1944).
- Lucas, A. A preliminary note on some preservative materials used by the ancient Egyptians in connection with embalming. *Cairo Sci. J.* **II**, 133–147 (1908).

12. Spielmann, P. E. To what extent did the ancient Egyptians employ bitumen for embalming? *J. Egypt. Archaeol.* **XVIII**, 177–180 (1932).
13. Zaki, A. & Iskander, Z. Materials and methods used for mummifying the body of Amentefnekht, Saqqara 1941. *Ann. Serv. Antiquites Egypte* **XLII**, 223–255 (1943).
14. Rullkötter, J. & Nissenbaum, A. Dead Sea asphalt in Egyptian mummies: molecular evidence. *Naturwissenschaften* **75**, 618–621 (1988).
15. Colombini, M. P., Modugno, F., Silvano, F. & Onor, M. Characterization of the balm of an Egyptian mummy from the seventh century B.C. *Stud. Conserv.* **45**, 19–29 (2000).
16. Koller, J., Baumer U., Kaup Y., Etspüler, H. & Weser, U. Embalming was used in Old Kingdom. *Nature* **391**, 343–344 (1998).
17. Buckley, S. A., Stott, A. W. & Evershed, R. P. Studies of organic residues from ancient Egyptian mummies using high temperature-gas chromatography-mass spectrometry and sequential thermal desorption-gas chromatography-mass spectrometry and pyrolysis-gas chromatography-mass spectrometry. *Analyst* **124**, 443–452 (1999).
18. Mills, J. S. & White, R. *The Organic Chemistry of Museum Objects* 2nd edn (Butterworth-Heinemann, Oxford, 1994).
19. Raven, M. J. Wax in Egyptian magic and symbolism. *Oudheidkundige Mededeelingen uit het Rijksmuseum van Oudheden te Leiden* **64**, 7–47 (1983).
20. van der Doelen, G. A., van den Berg, K. J. & Boon, J. J. Comparative chromatographic and mass spectrometric studies of triterpenoid varnishes: fresh material and aged samples from paintings. *Stud. Conserv.* **43**, 249–264 (1998).
21. Serpico, M. & White, R. in *Ancient Egyptian Materials and Technology* (eds Nicholson, P. T. & Shaw, I.) 430–474 (Cambridge Univ. Press, Cambridge, 2000).
22. Barber, M. S., McConnell, V. S. & DeCaux, B. S. Antimicrobial intermediates of the general phenylpropanoid and lignin specific pathways. *Phytochemistry* **54**, 53–56 (2000).
23. Bahn, P. G. The making of a mummy. *Nature* **356**, 109 (1992).
24. Connan, J. & Dessort, D. Du bitume dans des baumes de momies égyptiennes (1295 av. J.-C. -300 ap. J.-C.): détermination de son origine et évaluation de sa quantité. *C. R. Acad. Sci. Paris II* **312**, 1445–1452 (1991).
25. Granville, A. B. An essay on Egyptian mummies, with observations on the art of embalming amongst the ancient Egyptians. *Phil. Trans. R. Soc. Lond.* **115**, 269–316 (1825).

Acknowledgements

We thank S. Giles of Bristol Museum; J. Spencer and J. Taylor of the British Museum; J. Hayward of Liverpool Museum; R. David of Manchester Museum; and K. Eremin of the National Museum of Scotland for making the samples available to us. We thank J. Fletcher for advice on aspects of Egyptology; and J. Carter and A. Gledhill for technical assistance. NERC provided financial support for mass spectrometry facilities.

Correspondence and requests for materials should be addressed to R.P.E.

Local interactions predict large-scale pattern in empirically derived cellular automata

J. Timothy Wootton

Department of Ecology & Evolution, The University of Chicago, Chicago, Illinois 60637, USA

An important unanswered question in ecology is whether processes such as species interactions that occur at a local scale can generate large-scale patterns seen in nature^{1,2}. Because of the complexity of natural ecosystems, developing an adequate theoretical framework to scale up local processes has been challenging. Models of complex systems can produce a wide array of outcomes; therefore, model parameter values must be constrained by empirical information to usefully narrow the range of predicted behaviour. Under some conditions, spatially explicit models of locally interacting objects (for example, cells, sand grains, car drivers, or organisms), variously termed cellular automata^{3,4} or interacting particle models⁵, can self-organize to develop complex spatial and temporal patterning at larger scales in the absence of any externally imposed pattern^{1,6–8}. When these models are based on transition probabilities of moving between ecological states at a local level, relatively complex versions of these models can be linked readily to empirical information on ecosystem dynamics. Here, I show that an empirically derived cellular automaton

model of a rocky intertidal mussel bed based on local interactions correctly predicts large-scale spatial patterns observed in nature.

Mussel beds characterize many temperate, rocky intertidal shores throughout the world⁹, and can exhibit complex patterning. This patterning is thought to result from the interplay between interactions among sessile species and external disturbance agents, particularly large waves^{10–15}. Along the Pacific coast of North America, California mussels (*Mytilus californianus*) seem to interact locally with other organisms in two ways. First, mussels invade bare rock or displace species adjacent to them by overtopping their neighbours as they grow, and by creeping into adjacent areas by means of sequential attachment of new byssal threads in uncolonized areas as old byssal threads in colonized areas break or become detached¹². Second, because mussels attach to each other with byssal threads, when waves dislodge a mussel from the rock that mussel pulls its neighbours along with it, thereby transmitting the disturbance through the mussel bed in much the same way as forest fires propagate when sparks from a burning tree ignite neighbouring trees¹². Wave disturbance also acts differentially on mussel beds of different ages, with beds of old, multi-layered mussels being more prone to disturbance, owing to the high ratio of mussel mass per unit attachment area to the rocks, and to the high fraction of byssal threads attached to other mussels, rather than the rock^{12,14,15}. Other sessile species and their mobile associates colonize the bare rock created by disturbance events, and exhibit a complex set of interactions among themselves and with mussels^{10,12,13,16–19}. Because of its rapid dynamics and macroscopic organisms of modest size and mobility, the rocky intertidal zone is an unusually tractable model system in ecology for experimental investigations and for readily observing spatial and temporal dynamics in a complex natural ecosystem¹⁹. These features make mussel beds particularly suitable for the purpose of developing and testing different versions of models with empirical parameters, such as cellular automata.

Data on the dynamics of mussel beds were collected over a six-year period (May 1993 to May 1999) at 1,400 fixed points in the mussel bed of Tatoosh Island in Washington State²⁰. The data were taken from 14 quadrats (60 × 60 cm) positioned with fixed corner

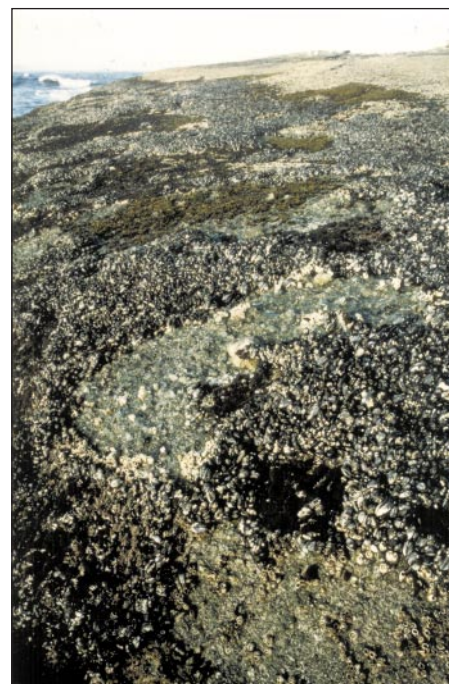


Figure 1 Photograph of spatial patterning in the mussel bed of Tatoosh Island, Washington.

pins, containing a 10×10 coordinate grid that was defined by monofilament lines spaced 6 cm apart. The ecological state (bare space, species identity, or size class of a species) of each point was determined at both a spring (May) and autumn (late August to early September) census each year. The occurrence or absence of a disturbance event after the previous census was noted for each point. These data were converted into transition probabilities and applied to a cellular automaton model of the mussel bed. The model was based on a 100×300 array of 6×6 cm cells with absorbing boundaries, which is similar in scale (6×18 m) to the average-sized mussel bed on individual rock benches in the intertidal zone of the study site¹². The full model included 15 ecological states (see Methods), substantially more than many cellular automata⁶. It also incorporated seasonal differences in transition and disturbance rates, and effects of having neighbouring mussels on transition rates (see Methods). The model ran for 500 simulated years, which minimized transient dynamics, and model output was collected on the cumulative average proportion of ecological states and on the distribution of gap sizes. Gap sizes were defined as the number of consecutive steps in random transect lines through the simulated mussel bed taken without encountering a cell filled with a mussel. These model predictions were tested against the size frequency distribution of gaps and the proportional composition of ecological states observed in independent transects taken in the mussel zone of the study site (see Methods).

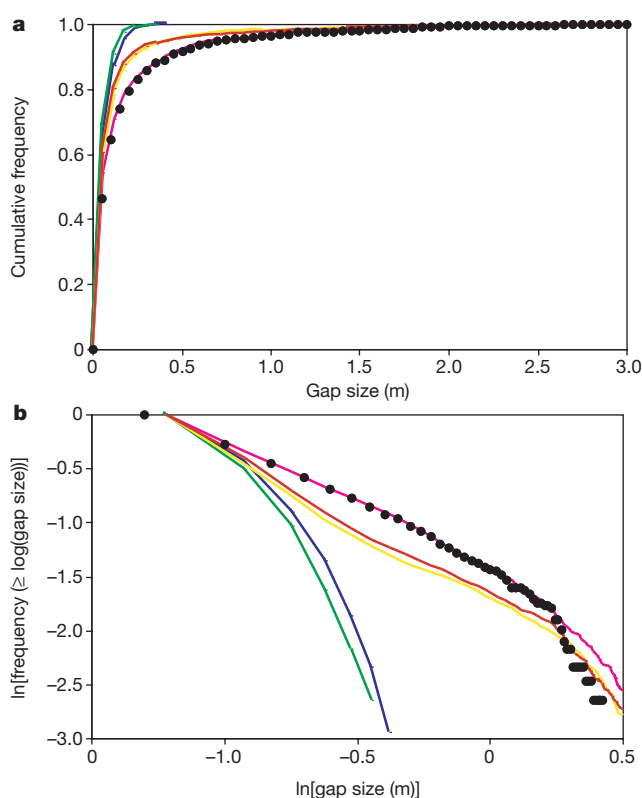


Figure 2 Frequency distribution of gap size (length of consecutive non-*M. californianus* points encountered in random transects) observed in mussel beds of Tatoosh Island, Washington (filled circles; $n = 876$), and generated by various versions of mussel-bed models that had empirical parameters. **a**, Standard cumulative frequency distribution of gap sizes. Pink line, local interactions and disturbance propagation ($P > 0.25$); purple line, non-spatial Markov model ($P < 0.00001$); green line, local interactions only ($P < 0.00001$); yellow line, mussels/no mussels, size structure ($P < 0.0001$); red line, forest fire model, mussel-bed parameters ($P < 0.0001$). **b**, Plot of log-reverse cumulative frequency (log frequency of gaps larger than the gap size in question) versus log gap size. Power law patterns appear as approximately linear curve segments in **b**. Colours are to the same scheme as **a**.

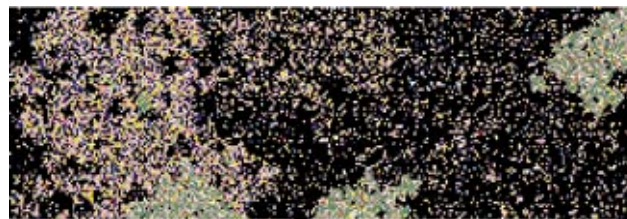


Figure 3 Example of spatial structure generated by the mussel-bed simulation model with external wave-initiated disturbance events. Black points are sites occupied by *M. californianus*. New gaps are dominated by rock (grey) and ephemeral algae (green) points. Older gaps with a variety of algal and invertebrate taxa are represented by different colours (for example: pink, *C. vancouveriensis*; blue, *M. trossulus*; white, *P. polymerus*).

Real mussel beds exhibit strong spatial patterning (Figs 1 and 2; $n = 876$). Results of the full model with empirical parameters self-organized to produce strong spatial patterning (Fig. 3). The distribution of gap sizes from random transects that were generated from the model followed a distribution similar to that observed from real mussel beds (Kolmogorov–Smirnov test, $P > 0.25$; Fig. 2). Additionally, the model accurately predicted the composition of ecological states observed in the study site (Fig. 4; 92.5% of the variance in composition correctly predicted).

Variants of the model were also explored to determine which processes were essential to the results of the model (see Box 1). Features of the full model that were altered included (1) eliminating all spatial structure; (2) removing explicit consideration of all taxa other than *M. californianus*; (3) eliminating the effects of mussel size structure; and (4) altering how disturbance was treated in the model by not explicitly considering the disturbance process. In all cases, the observed distribution of gap sizes differed significantly from model predictions (Kolmogorov–Smirnov tests, all $P < 0.0001$; Fig. 2). Therefore, details of species interactions, size structure, interactions among neighbours and the nature of disturbance impact and propagation were all essential in accurately predicting empirically observed spatial patterning.

In contrast to spatial patterning, the taxonomic composition of the community was accurately predicted by all model variants that included multiple taxa (Fig. 4), including a non-spatial Markov chain model²⁰, which predicted 98.4% of the variance in taxonomic composition. The high degree of predictive ability in this model suggests either that the observed spatial patterning is not an important determinant of the community composition of sessile

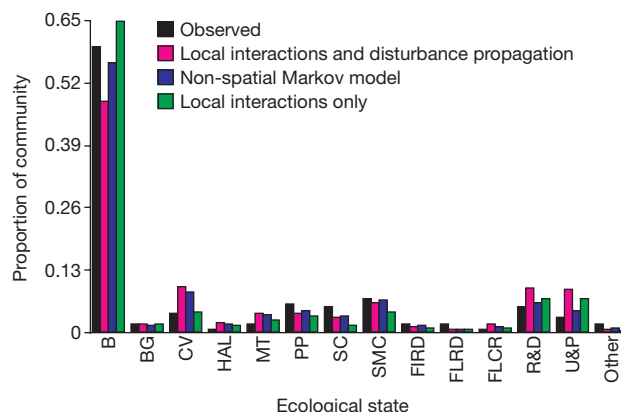


Figure 4 Composition of the mussel-bed community in each ecological state observed from transects (black bars) and predicted by different mussel-bed models with empirical parameters. B, large *M. californianus*; BG, *B. glandula*; CV, *C. vancouveriensis*; HAL, *H. glandiforme*; MT, *M. trossulus*; PP, *P. polymerus*; SC, *S. cariosus*; SMC, small *M. californianus*; FIRD, filamentous red algae; FLRD, foliose red algae; FLCR, fleshy crustose algae; R&D, rock and diatoms; U&P, ephemeral algae; Other, other taxa.

species, or that its effects are adequately subsumed into the parameter estimates of the non-spatial model. Furthermore, this result indicates that developing spatially explicit models that are data intensive may be unnecessary to understand and study compositional aspects of this system.

Model variants that differed in their treatment of disturbance (Box 1) provided some insight into the role of disturbance in pattern formation. A cellular automaton model variant in which transitions depended on neighbours, but which did not explicitly model disturbance, failed to produce patterning (Fig. 2), demonstrating that local interactions alone were insufficient to generate the observed, strong spatial patterning. In contrast, a model variant that lacked size structure and did not account explicitly for other species produced strong spatial patterning, but with a different distribution of gap sizes than was observed (Fig. 2; more small gaps and fewer very large gaps). This model, with parameters estimated from the mussel bed, is identical to the well studied forest fire model^{7,21,22}, for which self-organized critical behaviour has been reported. Another variant of the full-mussel-bed model that included all features of the full model except for other species generated a gap-size distribution very similar to the distribution of the forest fire model. Therefore, the details of interactions among other species within the community are primarily responsible for the deviations in the pattern of gap size between the full model and the forest fire model.

The results of this study demonstrate that empirically observed rates of species transitions and disturbance can produce self-organized behaviour in a cellular automaton model of the intertidal ecosystem. Furthermore, this self-organized patterning predicts well the patterning seen in nature. The results also indicate that details of size structure and species interactions are important in accurately generating natural patterning. The development of natural patterning only after incorporating external disturbance

propagation suggests, however, that conditions in the single-species forest fire models required to generate self-organized patterning, including low disturbance probability and markedly slower rates of recovery relative to disturbance propagation²¹, are important characteristics of the system.

A principal unanswered question in ecology is whether interactions at experimentally tractable local scales can be scaled up to recreate larger-scale patterns. Finding approaches that can successfully scale up local processes in complex systems is vital to ensure that detailed mechanistic studies are broadly relevant, and may be essential to fully understanding and predicting the impacts of environmental change¹. Furthermore, such approaches must be empirically accessible to be applied usefully. Using a cellular automaton framework and transition probabilities derived from a study of small-scale interactions (operating on an aerial scale of 36 cm²), I successfully reproduced patterning over an area that was more than four orders of magnitude larger. Therefore, it seems possible under some circumstances to derive insight into regional patterns from detailed studies of ecological processes at smaller, tractable scales. □

Methods

Ecological states included in the model

The 15 ecological states included bare space (with and without diatoms); three size/age classes of mussels (<2 cm wide (small mussel); >2 cm wide, newly occupying a site (new, large mussel); >2 cm wide occupying a site for more than one consecutive census (old, large mussel)); and 11 categories representing other common species (*Balanus glandula*, *Corallina vancouveriensis*, *Halosaccion glandiforme*, *Mytilus trossulus*, *Pollicipes polymerus*, *Semibalanus cariosus*) or clumped functional groups (filamentous red algae, foliose red algae, fleshy crustose algae, foliose ephemeral algae and all other taxa)²⁰. Rarer species were clumped into functional groups to ensure that sample sizes were sufficiently large to estimate transition probabilities.

Model simulation procedure

First, potential disturbance-initiation events were assigned randomly to single grid cells on the basis of previously reported numbers of disturbance events on rock benches of known area¹², divided by the observed proportion of points in the quadrats occupied by large, old mussels. Second, a disturbance was initiated at any cell to which a disturbance event was assigned if that cell contained an old large mussel. Third, for each newly disturbed cell that had been occupied by an old, large mussel, the model transmitted disturbance to the four nearest neighbours, on the basis of the observed probability that the disturbance was transmitted locally, which was derived from the quadrat data. The third step was repeated until no new disturbed cells were generated. Finally, transitions were randomly assigned to each cell on the basis of empirically derived transition probabilities that depended on the prior species identity of the cell, whether the cell was disturbed, and whether the cell previously had at least one mussel as a neighbour. The entire procedure was repeated alternately for parameters derived for summer (May–September) and winter (September–May).

Derivation of empirically observed patterns to test model

Observed composition and gap-size data were collected from transects that were randomly placed through the mussel bed at various spots throughout the study site. For composition measures, the ecological state under points at 0.5 m intervals were recorded ($n = 1,128$)²⁰. In a second series of transects measuring gap sizes, the presence or absence of mussels was noted at 5-cm intervals and gap size was defined as the number of sequential observations of points without mussels ($n = 876$ gaps). These data were tested against model predictions, which were best-fit functions for the cumulative gap size distributions generated by the model using nonlinear least squares regression. The best-fit functions were compared with the observed cumulative gap-size distribution using one-sample Kolmogorov–Smirnov tests²³.

Received 5 July; accepted 6 September 2001.

1. Levin, S. A. The problem of pattern and scale in ecology. *Ecology* **73**, 1943–1967 (1992).
2. Wiens, J. A., Stenseth, N. C., Van Horne, B. & Ims, R. A. Ecological mechanisms and landscape ecology. *Oikos* **66**, 369–380 (1993).
3. Wolfram, S. Cellular automata as models of complexity. *Nature* **311**, 419–424 (1984).
4. Ermentrout, G. B. & Edelstein-Keshet, L. Cellular automata approaches to biological modeling. *J. Theor. Biol.* **160**, 97–133 (1993).
5. Durrett, R. & Levin, S. A. Stochastic spatial models: a user's guide to ecological applications. *Phil. Trans. R. Soc. Lond. B* **343**, 329–350 (1994).
6. Bak, P. *How Nature Works: The Science of Self-organized Criticality* (Copernicus, New York, 1996).
7. Jensen, J. H. *Self-organized Criticality* (Cambridge Univ. Press, Cambridge, 1998).
8. Solé, R. V. & Manrubia, S. C. Are rainforests self-organized in a critical state? *J. Theor. Biol.* **173**, 31–40 (1995).
9. Stephenson, T. A. & Stephenson, A. *Life between Tidemarks on Rocky Shores* (Freeman, San Francisco, 1972).

Box 1

Summary of model variants examined and their essential differences

Full cellular automaton model. This model contained 15 ecological states, which accounted for species/taxonomic composition, empty space and mussel size/age structure. Disturbance was explicitly included by initiating disturbance at randomly assigned points and propagating the disturbance according to empirically observed rates of transmission. Transitions among ecological states were conditional on the season, the presence of mussel neighbours and whether a site was disturbed.

Markov chain model²⁰. Transitions among the 15 ecological states were not conditional on spatial arrangement. There was no dependency on neighbours and no explicit modelling of disturbance.

Local interactions model. Transitions among the 15 ecological states were conditional on the presence of neighbouring mussels. There was no explicit modelling of disturbance.

Basic mussel/no mussel model (forest fire model^{7,21,22}). For this model there were two ecological states (mussels, no mussels). There were no effects of interactions among other species or effects of mussel size/age class. Disturbance was explicitly modelled as in the full cellular automaton model. There were no seasonal effects.

Detailed mussel/no mussel model. This model included mussel size/age structure, seasonal effects and explicit disturbance as modelled in the full cellular automaton model. There were no effects of interactions among other species.

10. Dayton, P. K. Competition, disturbance, and community organization: the provision and subsequent utilization of space in a rocky intertidal community. *Ecol. Monogr.* **41**, 351–389 (1971).
11. Levin, S. A. & Paine, R. T. Disturbance, patch formation, and community structure. *Proc. Natl Acad. Sci. USA* **71**, 2744–2747 (1974).
12. Paine, R. T. & Levin, S. A. Intertidal landscapes: disturbance and the dynamics of pattern. *Ecol. Monogr.* **51**, 145–178 (1981).
13. Sousa, W. P. Intertidal mosaics: patch size, propagule availability and spatially variable patterns of succession. *Ecology* **65**, 1918–1935 (1984).
14. Denny, M. W. Lift as a mechanism of patch initiation in mussel beds. *J. Exp. Mar. Biol. Ecol.* **113**, 231–246 (1987).
15. Wootton, J. T. Size-dependent competition: effects on the dynamics vs. the end point of mussel bed succession. *Ecology* **74**, 195–206 (1993).
16. Johnson, L. E. Potential and peril of field experimentation: the use of copper to manipulate molluscan herbivores. *J. Exp. Mar. Biol. Ecol.* **160**, 251–262 (1992).
17. Wootton, J. T. Indirect effects, prey susceptibility, and habitat selection: impacts of birds on limpets and algae. *Ecology* **73**, 981–991 (1992).
18. Wootton, J. T. Predicting direct and indirect effects: an integrated approach using experiments and path analysis. *Ecology* **75**, 151–165 (1994).
19. Paine, R. T. *Marine Rocky Shores and Community Ecology: An Experimentalist's Perspective* (Ecology Institute, Olendorf, Germany, 1994).
20. Wootton, J. T. Prediction in complex communities: analysis of empirically-derived Markov models. *Ecology* **82**, 580–598 (2001).
21. Clar, S., Drossel, B. & Schwabl, F. Forest fires and other examples of self-organized criticality. *J. Phys. Condens. Matter* **8**, 6803–6824 (1996).
22. Malamud, B. D., Morein, G. & Turcotte, D. L. Forest fires: an example of self-organized critical behavior. *Science* **281**, 1840–1842 (1998).
23. Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, New York, 1995).

Acknowledgements

I thank the Makah Tribal Council and the United States Coast Guard for providing access to Tatoosh Island; A. Miller, R. Raynor, K. Rose, J. Salamunovitch, B. Scott, J. Sheridan, F. Stevens and A. Sun for field assistance; and P. Kareiva, R. Paine, C. Pfister and A. Sun for helpful comments. The work was supported in part by the A. W. Mellon Foundation and personal funds.

Correspondence and request for materials should be addressed to J.T.W. (e-mail: twootton@uchicago.edu).

Perceptual learning without perception

Takeo Watanabe*, José E. Náñez† & Yuka Sasaki‡

* Department of Psychology, Boston University, 64 Cummington Street, Boston, Massachusetts 02215, USA

† Department of Social and Behavioral Sciences, Arizona State University West, Phoenix, Arizona 85069-7100, USA

‡ NMR Center, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA

The brain is able to adapt rapidly and continually to the surrounding environment, becoming increasingly sensitive to important and frequently encountered stimuli^{1–4}. It is often claimed that this adaptive learning is highly task-specific, that is, we become more sensitive to the critical signals in the tasks we attend to^{5–15}. Here, we show a new type of perceptual learning, which occurs without attention, without awareness and without any task relevance. Subjects were repeatedly presented with a background motion signal so weak that its direction was not visible; the invisible motion was an irrelevant background to the central task that engaged the subject's attention. Despite being below the threshold of visibility and being irrelevant to the central task, the repetitive exposure improved performance specifically for the direction of the exposed motion when tested in a subsequent suprathreshold test. These results suggest that a frequently presented feature sensitizes the visual system merely owing to its frequency, not its relevance or salience.

We repeatedly presented a dynamic random dot (DRD) display¹⁶

with a small number of temporarily, coherently moving (signal) dots and a much larger number of randomly moving (noise) dots in the background. During this display, subjects conducted an identification task of letters that were presented at the centre of the display¹⁷ so that motion was an irrelevant feature. The ratio of signal to noise dots was a constant 5% during exposure. This signal strength was below the thresholds on tests of direction identification, direction discrimination and coherent motion detection tasks, which were conducted before, during and after the exposure. This suggests that attention and awareness had no access to the motion signal during the letter identification task. Nevertheless, after repeated exposure to this 'invisible', 5% level of coherent motion in the background, a significant performance improvement was obtained only for the motion directions within a small range around the exposed coherent direction when tested with a DRD display with 10% coherent motion. That is, this subsequent test revealed perceptual learning as a result of mere exposure to an 'invisible' motion direction.

The first experiment consisted of an exposure stage that was preceded and followed by control test stages (Fig. 1a). In each trial of

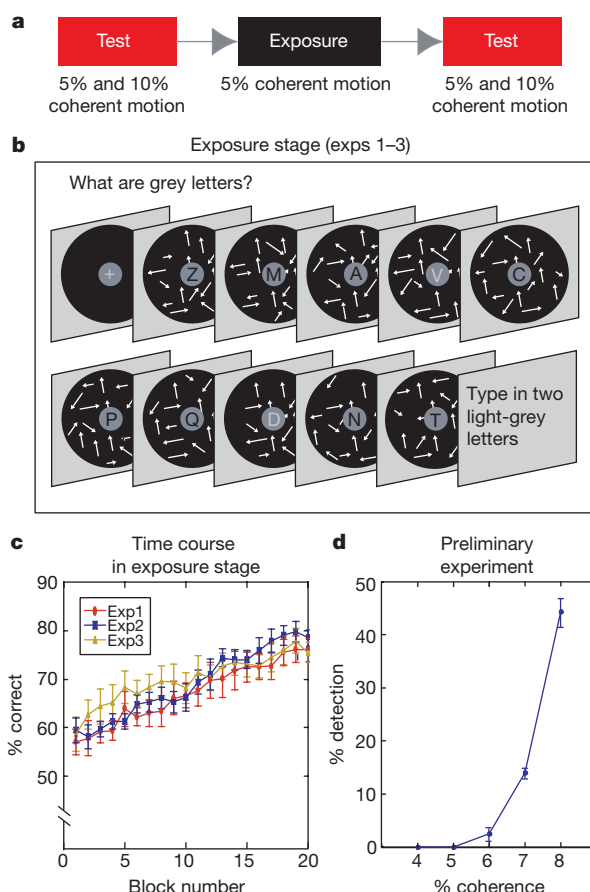


Figure 1 General procedure and exposure stage. **a**, In the exposure stage, dynamic random dot (DRD) displays with 5% coherent motion were presented; these were irrelevant to the task. In the test stages, DRD displays with 5% and 10% coherence were presented; these were relevant to the task. **b**, Exposure stage procedure. A sequence consisted of 8 black letters (0.5 cd m^{-2}) and 2 light-grey letters (53.0 cd m^{-2}) on a 1° diameter, dark-grey circle (32.0 cd m^{-2}), which was within the 10° diameter, black circular background (0.5 cd m^{-2}). Each letter was presented for 33 ms and was followed by a 17-ms blank interval. Thus, the duration of the sequence and the motion background was 500 ms. **c**, Time course of the mean 'correct' percentage ($\pm$ s.e.) for eight subjects in the exposure stage in experiments 1–3. **d**, Mean percentage ($\pm$ s.e. for eight subjects) of coherence detection as a function of coherence percentage in the preliminary experiment.

the exposure stage, ten letters were successively presented within a grey circle. During letter presentation, moving white dots were also presented on a surrounding black annulus background (Fig. 1b). Five per cent of the dots in a successive frame moved in the same direction while the remaining 95% of the dots were replaced, so that they were perceived to move randomly. This 5% coherence is lower than the detection threshold of 8.3% for coherent motion found in our preliminary experiment (see Methods and Fig. 1d). Direction of the coherent motion was determined randomly for each subject and was constant throughout the exposure stage. The subjects were instructed to identify two light-grey letters in order of their appearance in a rapidly changing stream of black letters presented at fixation, while ignoring the moving dots in the black background (Fig. 1b). To measure the effect of exposure on the performance for motion direction, the test stages were conducted before and after the exposure stage. In contrast to the exposure stage in which only 5% coherent motion was presented, in the test stages not only 5% but also 10% motion was presented to each subject. Whereas a coherent motion direction was constant for each subject in the exposure stage, in the test stages coherent motion direction was varied in eight

steps from trial to trial. In each trial, the subjects were instructed to indicate which one of the eight coherent motion directions they perceived (Fig. 2a). Eye movements were monitored in all of the trials in the exposure and test stages.

The performance of letter identification in the exposure stage showed steady increase for all of the subjects (Fig. 1c). The performance obtained when tested with the 5% level of coherent motion before and after the exposure stage did not indicate any significant difference from chance level (Fig. 2b). This suggests that the subjects did not perceive precisely the 5% level of coherent motion during the exposure stage. When tested with 10% coherent motion in the test stages, performance improvement was significantly greater only for the exposed direction and its vicinity after the exposure stage, compared with before. These results were consistent across all of the subjects. Furthermore, there was no significant correlation between detected eye movements and the direction of coherent motion during the exposure stage, excluding the possibility that eye movements in the exposed coherent motion direction somehow improved performance in the second test stage.

To test whether the first test stage influenced performance in the

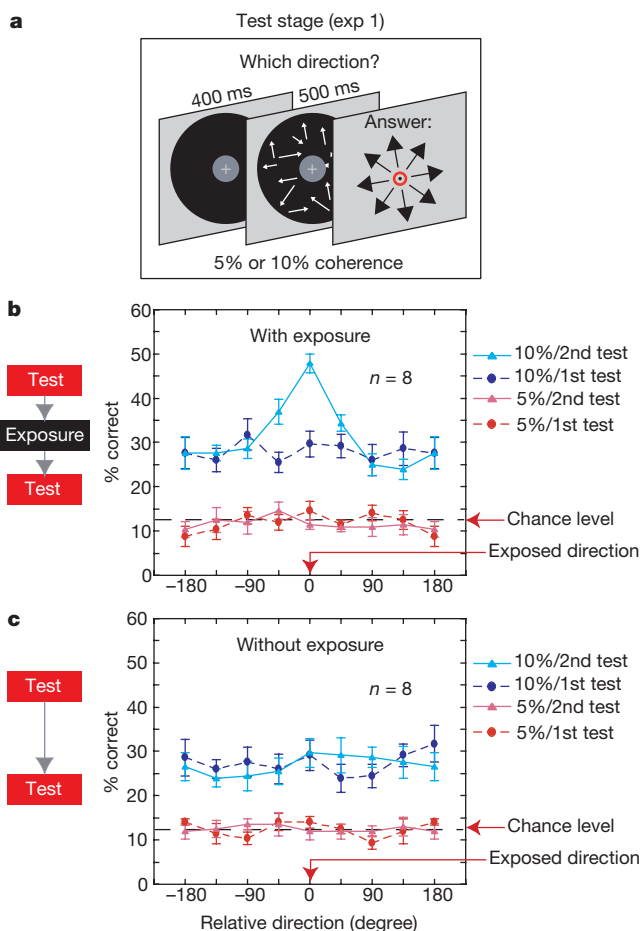


Figure 2 Test stages in experiment 1. **a**, Moving dots were replaced by eight arrows. Subjects placed a cursor on the arrow representing the perceived coherent direction. **b**, Mean correct percentage ($\pm$ s.e.) for eight subjects. A level of 5% coherence produced no significant difference between correct percentage and chance, for any direction, either for the main and control conditions combined in the first test stage, or for the main or control condition in the second test stage (Wilcoxon signed rank (WSR) test). For 10% coherence, only for the exposed direction ($P < 0.01$, WSR test) and for -45° ($P < 0.05$, WSR test), the percentage was significantly higher after exposure than before. The mean correct percentage for all of the directions was significantly higher than chance ($P < 0.01$, WSR test). **c**, No significant difference was found in correct percentage between the two test stages in the control condition without the exposure stage (WSR test).

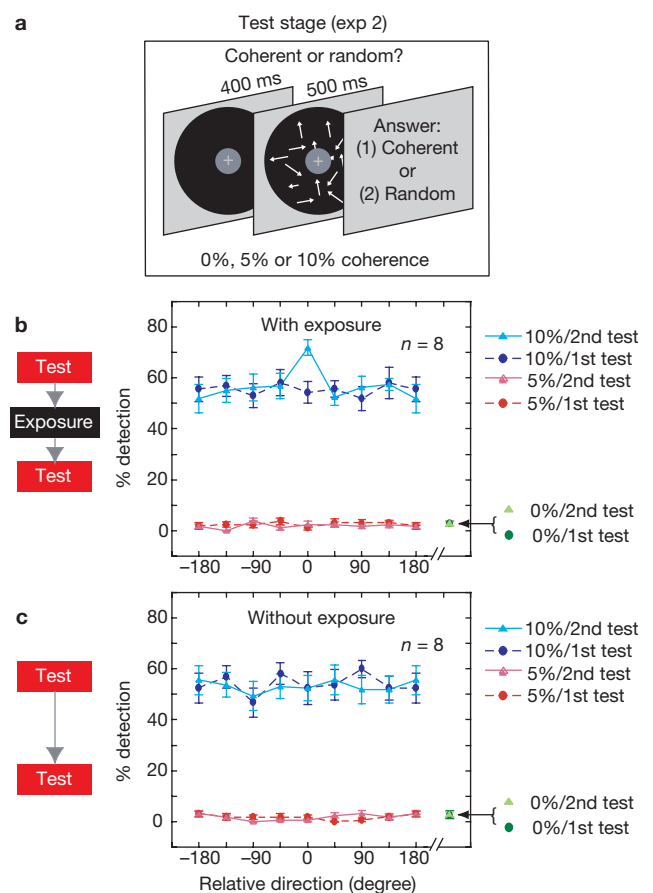


Figure 3 Test stages in experiment 2. **a**, After the disappearance of moving dots, subjects pressed a corresponding key to report whether coherent motion was perceived. **b**, Mean percentage ($\pm$ s.e. for eight subjects) for the detection of 0%, 5% and 10% coherent motion. No significant difference was found between 5% and 0% coherent motion, for any direction, either for the main and control conditions combined in the first test stage or for the main or control condition in the second test stage (WSR test). For 10% coherence, the percentage was significantly higher after the exposure than before, only for the exposed direction ($P < 0.05$, WSR test). The percentage for all of the directions was significantly higher than for 0% coherent motion ($P < 0.01$, WSR test). **c**, No significant difference was found in detection percentage between the two test stages in the control condition without the exposure stage (WSR test).

second, eight new subjects were tested in the two test stages only, without the exposure stage. No improvement was found either for the 5% or 10% coherent motion condition (Fig. 2c).

The overall results of the first experiment indicate that performance on a task specifying a coherent motion direction improved through mere exposure to the motion direction, which was sub-threshold and irrelevant to a task. To test whether the 5% coherent motion was actually invisible, we conducted a second experiment measuring the detectability of the coherent motion. The procedure for the exposure stage was identical to that used in the first experiment, and showed similar results (Fig. 1c). However, during the test stages in the second experiment, along with the 5% and 10% coherent motion displays, 0% (random motion) displays were also presented for each subject. The order of the presentation of these three kinds of displays was randomized from trial to trial. In each trial the subjects' task was to indicate whether they saw a group of dots moving coherently in the same direction (Fig. 3a).

The average likelihood of detecting the 5% coherent motion for any of the tested directions was nearly zero both before and after the exposure stage, and did not significantly differ from the average likelihood for the 0% condition (Fig. 3b). These results suggest that

the 5% level of coherent motion was invisible (undetectable) during the exposure stage. When tested with the 10% level of coherent motion, detectability was significantly higher after exposure compared with before, only in the exposed motion direction. The results were consistent across all subjects. No effect of a preceding test stage on a subsequent test stage was found in the control without exposure stage (Fig. 3c).

We next tested whether the ratio of the number of 0% coherence motion trials to the total number of trials influenced the subjects' internal criterion responses. A control experiment was conducted with eight new subjects in which the number of trials for the 0% motion coherence condition was increased to 240, whereas the number of trials for each of 5% and 10% motion coherence was kept at the original 160. We obtained similar results (not shown) to those in the main condition (Fig. 2b, c), indicating that irrespective of the difference in the ratio of the 0% coherent motion trials to the total number of trials, coherent motion was not perceived when tested with the 5% level of coherent motion.

The results of the second experiment suggest that repeated exposure to a 5% coherent motion direction, although below detection threshold, resulted in enhanced detectability of the previously exposed motion direction when observed at a higher, more detectable 10% level of coherent motion.

In the first experiment, the difference between the closest motion directions represented by the arrows in the test stages was 45°. This raises the question of whether mere exposure to the 5% level of coherent motion enhances finer-scaled discriminability of the exposed coherent direction. Thus, in a third experiment, we tested discriminability of coherent motion directions that differed only by 3°. The procedure of the exposure stage was identical to that of the first experiment and showed similar results (Fig. 1c). In the test stages, the subjects were instructed to indicate whether the direction of coherently moving dots in two successively presented displays matched (Fig. 4a). In a 'matching trial' condition, the coherent motion directions in the two displays matched. In a 'non-matching trial' condition, the directions differed only by 3° (ref. 7). The same tendency as in the first experiment was observed (Fig. 4b). The lack of a significant difference between the subjects' performance and chance level in test stages before and after the exposure stage suggests that the 5% coherent motion was invisible (non-discriminable) during the exposure stage. The performance increase in discrimination of the previously exposed motion direction from those differing by only $\pm 3^\circ$ when tested with 10% motion coherence suggests that mere exposure to the 5% invisible, coherent motion may fine tune the brain to the exposed motion direction. No effect of a preceding test stage on a subsequent test stage was found in the control without exposure stage (Fig. 4c).

In the above experiments, we found that the 5% level of coherent motion was neither discriminable nor detectable both before and after exposure to the motion. To further assure that subjects were unable to either detect or discriminate the 5% level of coherent motion while being exposed to it, in addition to the letter identification task the subjects performed a direction indication task (fourth experiment) or a coherent motion detection task (fifth experiment) in each trial of the exposure stage. Performance in the coherent motion indication task was near chance (12.5%) in all of the blocks of the exposure stages. The percentage of subjects detecting coherent motion was also not significantly different from 0% in any of the blocks of the exposure stage (not shown). These results further confirm that 5% coherent motion was indeed invisible throughout the exposure stage.

In all five experiments, with 5% coherent motion, performance in the direction indication, coherent motion detection and direction discrimination tasks in the test stages were all at chance level before, during and after the exposure phase. This provides compelling evidence that the 5% level of coherent motion was neither detectable nor discriminable, suggesting that attention and awareness had no

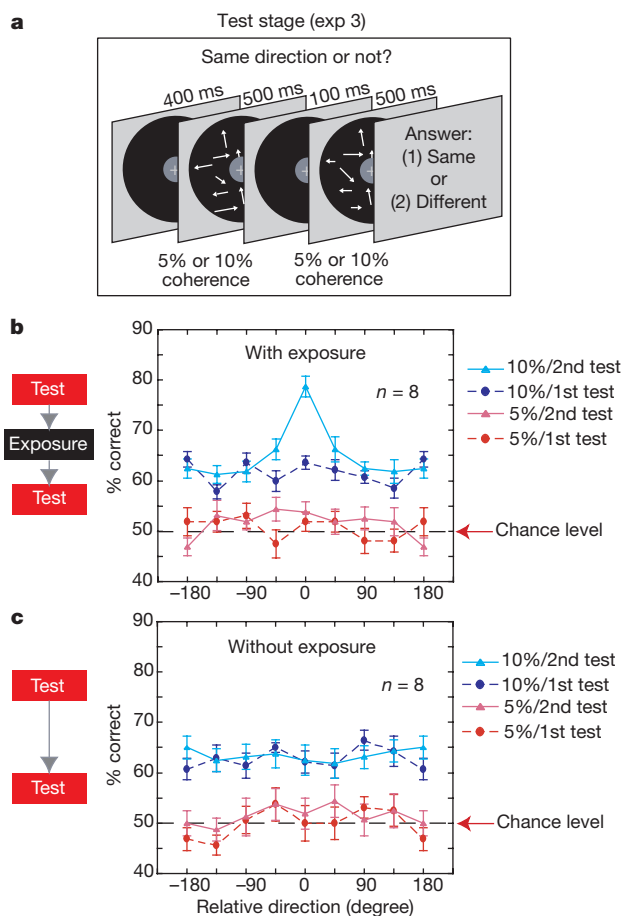


Figure 4 Test stages in experiment 3. **a**, Two DRD displays were presented, each for 500 ms with a 100-ms blank interval. After the disappearance of the second display, subjects pressed a corresponding key to report whether the coherent directions in the displays matched. **b**, Mean correct percentage ($\pm$ s.e. for eight subjects) for 5% and 10% coherent motion, before and after the exposure. For 5% coherence, the same statistical results were obtained as in experiment 1. For 10% coherence, the percentage was significantly higher after the exposure than before for the exposed direction only ($P < 0.01$, WSR test). The percentage for all of the directions was significantly higher than chance ($P < 0.01$, WSR test). **c**, No significant difference was found in correct percentage between the two test stages in the control condition without the exposure stage (WSR test).

access to the motion signal during the letter identification task. Nevertheless, mere exposure to the 5% level of invisible, coherent motion resulted in enhanced performance specific for the exposed direction when subjects were subsequently tested with more visible (10%) coherent signals in all of the direction indication, coherent motion detection and direction discrimination tasks. We conclude that perceptual learning and modification of the underlying sensory mechanism specific for the exposed motion direction results from mere exposure to invisible, but physically present, motion.

Our findings have important implications for the question of how the brain adjusts its visual capabilities to the environment. It seems that frequent presentation of a feature, such as a particular direction of motion, sensitizes the visual system to that feature, even if the feature is perceptually invisible, and even if it is irrelevant to a given task. Although the present finding does not deny the important role of attention in perceptual learning^{8,11,13–15} and in motion^{18–22}, it indicates that the adult brain has the flexibility to adapt to certain features of the environment as a result of mere exposure, even if the feature is so weak that it does not lead to awareness. This unsupervised modification would seem to have unwanted consequences: when we are bombarded by unimportant information we hope that by consciously ignoring it, we will escape its effects. This belief now seems unwarranted. The brain seems to equate frequency with ecological importance^{23–28}, and even though this may once have been a good strategy, it may be less adaptive in the context of our manipulative, modern-day media. It is possible that the present finding might be a result of unsupervised learning applied to lateral inhibition between neurons for different directions²⁹. □

Methods

Subjects

The age of the subjects ranged between 19 and 25 years. All of the subjects had normal or corrected-to-normal vision and were naive to the purpose of the experiments. For each of the main and control conditions of all experiments, eight (experiments 1–3) or six (experiments 4–5) new subjects were used.

Preliminary threshold measurement

The 8.3% coherence motion detection threshold with the DRD display used in the test stages was established as a result of a preliminary experiment with the staircase method ($n = 6$). In the method, each time the subject's response (visible or invisible) changed, the direction of changes in signal-to-noise ratio (smaller or larger) is altered. The 10% coherent motion display was chosen to be used in the main experiments because its signal-to-noise ratio is over the established 8.3% threshold. It is uncertain whether subjects at no time perceived the coherent motion at the threshold. To determine the signal-to-noise ratio at which the subjects were unlikely to see the coherent motion, we asked eight new observers to report whether they saw coherent motion during 20 trials at each of 8, 7, 6, 5 and 4% coherence. None of the subjects reported any detection in the 5% and 4% coherence conditions (see Fig. 1d). Thus, we determined to use 5% coherence in the main experiments.

Design of the first experiment

The local dot speed was 14.2° per second in the motion displays. A non-optimal speed for detection of the coherent motion was chosen to ensure that detection would be poorer. The optimal speed, at least for a monkey, is known to be 3° per second¹⁶. Dot density was 1.27 dots per degree². Signal dots were randomly chosen in each frame. For example, for the 5% coherent motion display, 5% of the dots carried the signal from one frame to the next and then a different set of dots carried signals in the next frame transition¹⁶. Thus, the lifetime of almost all dots was the same as the frame duration of 50 ms.

For exposure stages, the first and second light-grey letters were presented in one of the first five serial positions and in one of the second five serial positions, respectively. They were determined randomly in each trial. Each block consisted of 960 trials and was completed on a different day. The exposure stage was terminated after 20 blocks (this was the number around which preliminary results indicated improvement asymptotes). No accuracy feedback was given to the subjects.

During the test stages each subject viewed the eight directions (−135°, −90°, −45°, 0°, 45°, 90°, 135° and 180°) from the exposed coherent direction, where positive and negative values represent clockwise and anticlockwise rotations. A trial for each direction was repeated 20 times. The order of the presentations of these conditions was randomly determined for each subject. No accuracy feedback was given to the subjects. The first test stage was conducted at least one day before the beginning of the exposure stage, and the second stage at least one day after the end of the exposure stage.

For measurements of eye movement, we used the ViewPoint Eye Tracker version 3

(Arrington Research) with approximately 0.3° resolution. For the control experiment with no exposure, subjects were given the same time interval between the two test stages (25 days), which was the mean interval in the main experiment (with exposure). The procedure of the test stages was identical to that of the main experiment.

Second experiment

Before the onset of the experiment, subjects were told that the number of trials in which coherent motion would be seen is not necessarily the same as that in the 0% coherent motion trials. Twenty trials were presented for each of the eight directions in the 5% and 10% coherent motion conditions, and 160 trials with the 0% coherent motion condition.

Third experiment

The test stages consisted of 50% matching trials and 50% non-matching trials. In the non-matching trials one of the eight directions (45° interval between the closest directions) was presented in the first display and the direction differing by $\pm 3^\circ$ from that was presented in the second display.

Fourth and fifth experiments

In each trial of the exposure stage of the fourth experiment, both the letter identification task and the direction indication task, which was the same as in the test stages of the first experiment, were conducted. After the offset of a sequence of ten letters, three subjects were instructed first to conduct a direction identification task and then a letter identification task. The order of the tasks was reversed for the remaining three subjects. In the fifth experiment, instead of the direction indication task as in the fourth experiment, subjects performed the coherent motion detection task as in the test stage of the second experiment (Fig. 4a).

Received 19 June; accepted 4 September 2001.

- Gilbert, C. D. Plasticity in visual perception and physiology. *Curr. Opin. Neurobiol.* **6**, 269–274 (1996).
- Vaina, L. M., Belliveau, J. W., des Roziers, E. B. & Zeffiro, T. A. Neural systems underlying learning and representation of global motion. *Proc. Natl Acad. Sci. USA* **95**, 12657–12662 (1998).
- Grossberg, S. How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spat. Vis.* **12**, 163–185 (1999).
- Zohary, E., Celebrini, S., Britten, K. H. & Newsome, W. T. Neuronal plasticity that underlies improvement in perceptual performance. *Science* **263**, 1289–1292 (1994).
- Ramachandran, V. S. & Braddick, O. Orientation-specific learning in stereopsis. *Perception* **2**, 371–376 (1973).
- Fiorentini, A. & Berardi, N. Perceptual learning specific for orientation and spatial frequency. *Nature* **287**, 43–44 (1980).
- Ball, K. & Sekuler, R. Direction-specific improvement in motion discrimination. *Vision Res.* **27**, 953–965 (1987).
- Shiu, L. P. & Pashler, H. Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept. Psychophys.* **52**, 582–588 (1992).
- Karni, A. & Sagiv, D. The time course of learning a visual skill. *Nature* **365**, 250–252 (1993).
- Fahle, M. & Edelman, S. Long-term learning in vernier acuity: effects of stimulus orientation, range and of feedback. *Vision Res.* **33**, 397–412 (1993).
- Ahissar, M. & Hochstein, S. Attentional control of early perceptual learning. *Proc. Natl Acad. Sci. USA* **90**, 5718–5722 (1993).
- Sagiv, D. & Tanne, D. Perceptual learning: learning to see. *Curr. Opin. Neurobiol.* **4**, 195–199 (1994).
- Ito, M., Westheimer, G. & Gilbert, C. D. Attention and perceptual learning modulate contextual influences on visual perception. *Neuron* **20**, 1191–1197 (1998).
- Crist, R. E., Li, W. & Gilbert, C. D. Learning to see: experience and attention in primary visual cortex. *Nature Neurosci.* **4**, 519–525 (2001).
- Koyama, S. & Watanabe, T. Different mechanisms for the learning of motion detection vs. the learning of motion direction discrimination. *J. Vis. (Suppl.)* (in the press).
- Newsome, W. T. & Pare, E. B. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 2201–2211 (1988).
- Joseph, J. S., Chun, M. M. & Nakayama, K. Attentional requirements in a 'preattentive' feature search task. *Nature* **387**, 805–807 (1997).
- Treue, S. & Maunsell, J. H. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).
- Watanabe, T. et al. Task-dependent influences of attention on the activation of human primary visual cortex. *Proc. Natl Acad. Sci. USA* **95**, 11489–11492 (1998).
- Watanabe, T. et al. Attention-regulated activity in human primary visual cortex. *J. Neurophysiol.* **79**, 2218–2221 (1998).
- Somers, D. C., Dale, A. M., Seiffert, A. E. & Tootell, R. B. Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl Acad. Sci. USA* **96**, 1663–1668 (1999).
- Gandhi, S. P., Heeger, D. J. & Boynton, G. M. Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl Acad. Sci. USA* **96**, 3314–3319 (1999).
- Nakayama, K. & Shimojo, S. Experiencing and perceiving visual surfaces. *Science* **257**, 1357–1363 (1992).
- Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
- Albert, M. K. & Hoffman, D. D. The generic-viewpoint assumption and illusory contours. *Perception* **29**, 303–312 (2000).
- Slemmer, J. A., Kirkham, N. G. & Johnson, S. P. Visual statistical learning in infancy. *J. Vis. (Suppl.)* (in the press).

27. Bar, M. & Biederman, I. Subliminal visual priming. *Psychol. Sci.* **9**, 464–469 (1998).
 28. Chun, M. M. Contextual cueing of visual attention. *Trends Cogn. Sci.* **4**, 170–178 (2000).
 29. Barlow, H. B. in *Vision: Coding and Efficiency* (ed. Blakemore, C.) (Cambridge Univ. Press, Cambridge, 1989).

Acknowledgements

We thank M. Ahissar, P. Cavanagh, S. Hochstein, W. T. Newsome, R. Raizada and R. Sekuler for their comments on this or related study. This work was supported by an NSF (SBES, HCP) grant to T.W. and by Chiba University and JSPS fellowship to Y.S.

Correspondence and requests for materials should be addressed to T.W.
 (e-mail: takeo@bu.edu).

Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18

J. Parkhill*, G. Dougan†, K. D. James*, N. R. Thomson*, D. Pickard†, J. Wain†, C. Churcher*, K. L. Mungall*, S. D. Bentley*, M. T. G. Holden*, M. Sebaihia*, S. Baker*, D. Basham*, K. Brooks*, T. Chillingworth*, P. Connerton†, A. Cronin*, P. Davis*, R. M. Davies*, L. Dowd*, N. White‡, J. Farrar‡, T. Feltwell*, N. Hamlin*, A. Haque†, T. T. Hien§, S. Holroyd*, K. Jagels*, A. Krogh||, T. S. Larsen||, S. Leather*, S. Moule*, P. Ó'Gaora†, C. Parry§, M. Quail*, K. Rutherford*, M. Simmonds*, J. Skelton*, K. Stevens*, S. Whitehead* & B. G. Barrell*

* The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

† Centre for Molecular Microbiology and Infection, Department of Biological Sciences, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK

‡ University of Oxford-Wellcome Trust Clinical Research Unit, Centre for Tropical Diseases, 190 Ben Ham Thu, Quan 5, Ho Chi Minh City, Vietnam, and Centre for Tropical Medicine, University of Oxford, OX3 9DU, UK

§ The Centre for Tropical Diseases, 190 Ben Ham Thu, Quan 5, Ho Chi Minh City, Vietnam

|| Centre for Biological Sequence Analysis, BioCentrum-DTU, The Technical University of Denmark, Building 208, 2800 Lyngby, Denmark

Salmonella enterica serovar Typhi (*S. typhi*) is the aetiological agent of typhoid fever, a serious invasive bacterial disease of humans with an annual global burden of approximately 16 million cases, leading to 600,000 fatalities¹. Many *S. enterica* serovars actively invade the mucosal surface of the intestine but are normally contained in healthy individuals by the local immune defence mechanisms. However, *S. typhi* has evolved the ability to spread to the deeper tissues of humans, including liver, spleen and bone marrow. Here we have sequenced the 4,809,037-base pair (bp) genome of a *S. typhi* (CT18) that is resistant to multiple drugs, revealing the presence of hundreds of insertions and deletions compared with the *Escherichia coli* genome, ranging in size from single genes to large islands. Notably, the genome sequence identifies over two hundred pseudogenes, several corresponding to genes that are known to contribute to virulence in *Salmonella typhimurium*. This genetic degradation may contribute to the human-restricted host range for *S. typhi*. CT18 harbours a 218,150-bp multiple-drug-resistance *incH1* plasmid (pHCM1), and a 106,516-bp cryptic plasmid (pHCM2), which shows recent common ancestry with a virulence plasmid of *Yersinia pestis*.

Salmonella typhi is a serovar of *S. enterica*, which is serologically O (lipopolysaccharide) type 09, 012; H (flagellin) type d; and Vi

(extracellular capsule) positive. Humans are the only known natural host of *S. typhi*, with *S. typhi* showing limited pathogenicity for most animals. Isoenzyme analysis has suggested that isolates of *S. typhi* around the world are highly related², a view confirmed using multi-locus sequence analysis (C. Kidgell *et al.*, unpublished data). Multiple drug resistance (MDR) is a serious emerging threat to the treatment of infectious diseases—MDR *S. typhi* are resistant to commonly available antibiotics, and clinical resistance to fluoroquinolones, the most effective antimicrobials for the treatment of typhoid fever, has been reported³. *Salmonella typhi* CT18 is an example of an emerging MDR microorganism; in depth genome analysis will contribute to our understanding of how such microorganisms adapt rapidly to new environmental opportunities that are presented by modern human society.

The principal features of the *S. typhi* CT18 chromosome and the two plasmids harboured by this strain are shown in Table 1. The beginning of the sequence was taken to correspond with minute 0 on the *E. coli* and *Salmonella* genetic maps⁴; the origin and terminus of replication, predicted by comparison with *E. coli* and confirmed by GC-bias (Fig. 1), are near 3.765 megabases (Mb) and 1.437 Mb, respectively. The metabolism of *Salmonella* and *E. coli* has been extensively studied over many decades⁴, and our analysis reveals few surprises in this area. However, the chromosome was predicted to encode 204 pseudogenes, which is remarkable in a genome of an organism capable of growth both inside and outside the host. Most of these pseudogenes (124 out of 204) have been inactivated by the introduction of a single frameshift or stop codon, which suggests that they are of recent origin. Five pseudogenes (*priC*, *ushA*, *fepE*, *sopE2* and *fliB*) were re-sequenced from several independent *S. typhi* isolates, and were identical in every case. Frameshifts that are due to changes in the length of homopolymeric tracts account for 45 pseudogenes; this is a mechanism of variation that was previously shown to occur in *E. coli*⁵, although at much lower rates than are required for rapid phase variation in other organisms. Some of the pseudogenes (27 out of 204) are the remnants of insertion sequence (IS) transposases, integrases and genes of bacteriophage origin. However, there are a significant number (75 out of 204) that are predicted to be involved in housekeeping functions, such as a component of the DNA primase complex, *priC*, cobalamin biosynthesis genes *chiM*, *-J*, *-K* and *-C*, the proline transporter *proV*, and the anaerobic dimethyl sulphoxide reductase components *dmsA* and *dmsB*. Many more mutations (46 out of 204) are in genes that are potentially involved in virulence or host interaction. Examples of this latter group include components of seven of the twelve

Table 1 Features of the *S. typhi* genome

Component of genome	Property
Chromosome	
Total size	4,809,037 bp
G+C content	52.09%
Coding sequences	4,599
... of which pseudogenes	204
Coding density	87.6%
Average gene length	958 bp
Ribosomal RNAs	6 × (16S–23S–5S), 1 × (16S–23S–5S–5S)
Transfer RNAs	78
Other stable RNAs	8
pHCM1	
Total size	218,150 bp
G+C content	47.58%
Coding sequences	249
... of which pseudogenes	8
Coding density	83.8%
Average gene length	759 bp
pHCM2	
Total size	106,516 bp
G+C content	50.6%
Coding sequences	131
... of which pseudogenes	0
Coding density	87.1%
Average gene length	708 bp
Transfer RNAs	1

chaperone–usher fimbrial operons⁶; the gene responsible for flagellar methylation, *fliB*; genes within or associated with previously described *Salmonella* pathogenicity islands (SPI) (for example, the sensor kinase *ttrS* that is associated with SPI-2 (ref. 7) and *cigR*, *marT* and *misL* from SPI-3 (ref. 8); the leucine-rich repeat protein *slrP*, which is involved in host-range specificity⁹ and is secreted through a type III system; other type-III-secreted effector proteins including *sseJ* (ref. 10), *sopE2* (ref. 11) and *sopA*; and the genes *shdA*, *ratA* and *sivH*, which are present in an island unique to *Salmonellae* infecting warm-blooded vertebrates¹². A greater proportion of pseudogenes than expected lie within islands that are unique to *S. typhi* compared with *E. coli* (59% (120 out of 204) compared with 33% (1,505 out of 4,599) of all genes). This apparent inactivation of

many of the mechanisms of host interaction may go some way to explaining the host restriction of *S. typhi* relative to other *Salmonella* serovars, and suggests that *S. typhi* may have passed through a recent evolutionary bottleneck.

Apart from some previously detected rearrangements¹³, which were caused by recombination between ribosomal RNA operons near the origin of replication, the genomes of *S. typhi* and *E. coli*¹⁴ are essentially collinear along their entire length (see Supplementary Information for a linear gene map and functional classification of identified genes; see also http://www.sanger.ac.uk/Projects/S_typhi). Although there are some cases of translocation of small gene blocks, most of the differences are due to insertions, deletions or replacements. Several of the large insertions that are present in

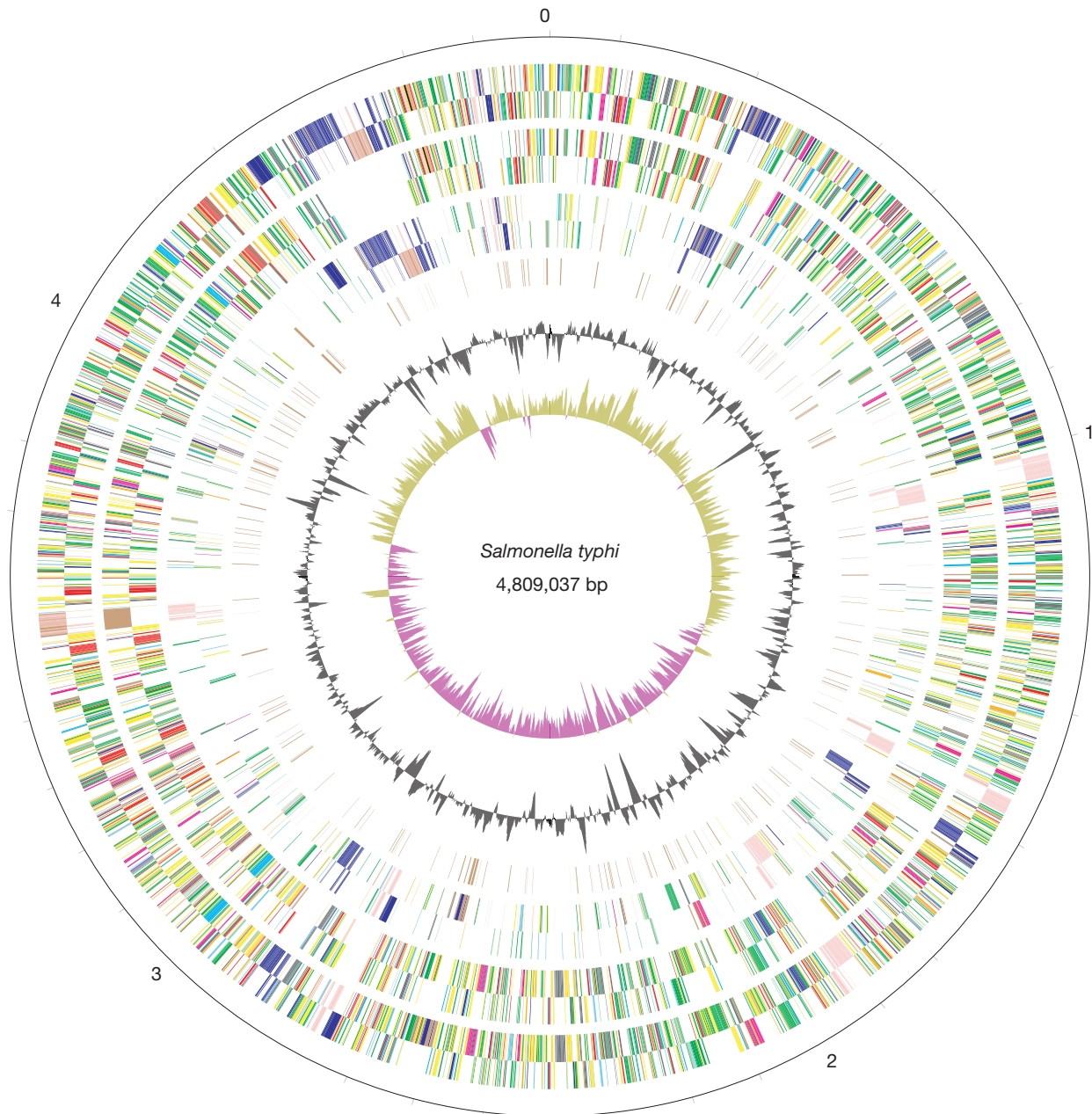


Figure 1 Circular representation of the *S. typhi* genome. The outer scale is marked in megabases. Circles range from 1 (outer circle) to 9 (inner circle). Circles 1 and 2, genes on forward and reverse strand; circles 3 and 4, genes conserved with *E. coli*; circles 5 and 6, genes unique to *S. typhi* with respect to *E. coli*; circle 7, pseudogenes; circle 8, G+C content; circle 9, GC bias ((G–C)/(G+C)); khaki indicates values >1; purple <1). All

genes are colour-coded by function: dark blue, pathogenicity/adaptation; black, energy metabolism; red, information transfer; dark green, membranes/surface structures; cyan, degradation of macromolecules; purple, degradation of small molecules; yellow, central/intermediary metabolism; light blue, regulators; pink, phage/IS elements; orange, conserved hypothetical; pale green, unknown function; brown, pseudogenes.

Salmonella have been extensively studied. These insertions generally carry genes that are important for survival in the host (including two type III secretion systems and an array of effector proteins and metal-ion transporters), they have often been inserted adjacent to a stable RNA gene, and may carry a gene encoding an integrase or transposase-like protein¹⁵. These regions, termed SPIs, are believed to be fairly recent horizontal acquisitions, and may be self-mobile. However, as observed in the comparison between *E. coli* K12 and 0157:H7 (ref. 16), there are many more insertions and deletions, including much smaller blocks (Fig. 2). Together there are 1,505 genes (32.7%) in 290 blocks that are unique to *S. typhi* relative to *E. coli* (see Supplementary Information), and 1,220 genes (28.4%) in 268 blocks that are unique to *E. coli* relative to *S. typhi*. Single-gene insertions account for 128 of the unique genes in *S. typhi*, and 456 genes are in insertions of 5 genes or less.

Among the larger blocks are the previously characterized SPIs 1–5 and seven prophage elements. There are also at least five more islands that have the characteristics of SPIs. Of the previously published SPIs we found that the original sequence of SPI-4 (ref. 17) varied most markedly from our sequence. SPI-4 was originally predicted to encode 18 genes, but our analysis of this region revealed the presence of only 8 coding sequences (CDS), two of which (STY4458 and STY4459) spanned most of those originally observed (these are represented by just one CDS in *S. typhimurium*). SPI-4 carries three CDS that are predicted to encode a type I secretion system—STY4458 and STY4459 are large, highly repetitive, and are weakly similar to RTX-like toxins. Of the new SPIs, SPI-6 (59 kb) encodes the *safA-D* and *tcsA-R* chaperone–usher fimbrial operons⁶. SPI-7 (134 kb) encodes the Vi biosynthetic genes¹⁸, the *SopE* prophage¹⁹ and a type IVB pilus operon²⁰. SPI-8 (6.8 kb) encodes two bacteriocin pseudogenes (STY3280 and STY3282) and a degenerate integrase; notably, genes conferring immunity to the bacteriocins remain intact. SPI-9 (16 kb), like SPI-4, encodes a type I secretory apparatus and a single, large RTX-like protein (STY2875). SPI-10 (33 kb) carries phage 46 and the *sefA-R* chaperone–usher fimbrial operon. In addition to these large islands, there are many insertions of smaller gene blocks and individual genes that may be involved in pathogenicity. These include: numerous secreted and integral membrane proteins; many new regulators; a set of eight genes that are potentially involved in extracellular polysaccharide biosynthesis (STY0759–0768, three of which are pseudogenes); a predicted iron-uptake ABC transporter (STY0802–0803); two puta-

tive efflux pumps (STY0278 and STY0414); and genes encoding several secreted effector proteins, including SifA, SopD, SopD2 and SspH2.

As would be expected, the relationship between *S. typhi* and *S. typhimurium*²¹ is very much closer than that between *S. typhi* and *E. coli*, although there are still significant differences. The same conservation of gene order is apparent (see Supplementary Information), disrupted only by the rearrangements around the rRNAs and a large inversion around the terminus in *S. typhimurium*¹³. As with *E. coli*, the differences are not limited to a few large blocks. Single-gene insertions account for 42 unique genes in *S. typhi*, and 103 genes are in insertions of 5 genes or less. In all, there are 601 genes (13.1%) in 82 blocks that are unique to *S. typhi* compared with *S. typhimurium* (see Supplementary Information), and 479 genes (10.9%) in 80 blocks that are unique to *S. typhimurium* relative to *S. typhi*. Several significant insertions are apparent in *S. typhi*. These insertions include: the *staA-G*, *tcfA-D*, *steA-G* and *stgA-G* chaperone–usher fimbrial systems⁶; a homologue of the *E. coli* haemolysin E (STY1498); a short insert carrying homologues of the *Campylobacter* toxin *cdtB* (STY1886) and the *Bordetella pertussis* toxin genes *ptxA* and *ptxB* (STY1890 and STY1891); a putative polysaccharide acetyltransferase (STY2629); phages 10, 15, 18 and 46; and SPIs 7, 8 and 10. Aside from these relatively few insertions, much of the difference in phenotype and host range between *S. typhi* and *S. typhimurium* may well be explained by the pseudogenes described above. Of the 204 pseudogenes that have been discovered in *S. typhi*, 145 are present as intact genes in *S. typhimurium*, whereas only 23 are present as pseudogenes (15 of which have the same inactivating mutations).

Salmonella typhi CT18 harbours two plasmids, the larger of which (pHCM1) is conjugative and encodes resistances to all of the first-line drugs used for the treatment of typhoid fever. pHCM1 shares around 168 kb of DNA at greater than 99% sequence identity with R27 (ref. 22)—an *incH1* plasmid that was first isolated in the 1960s from *S. enterica*. Shared functions include plasmid transfer and partition, and pHCM1 has apparently been derived by the insertion of 46 CDS, totalling around 50 kb, primarily at two positions in an R27-like ancestral plasmid (Fig. 3). Eighteen of these CDS are involved with resistance to antimicrobial agents or heavy metals, and 16 are of unknown function—just three of which are similar to other plasmid-borne genes. Several intact and degenerate integrases and transposases are clustered around these two regions of insertion,

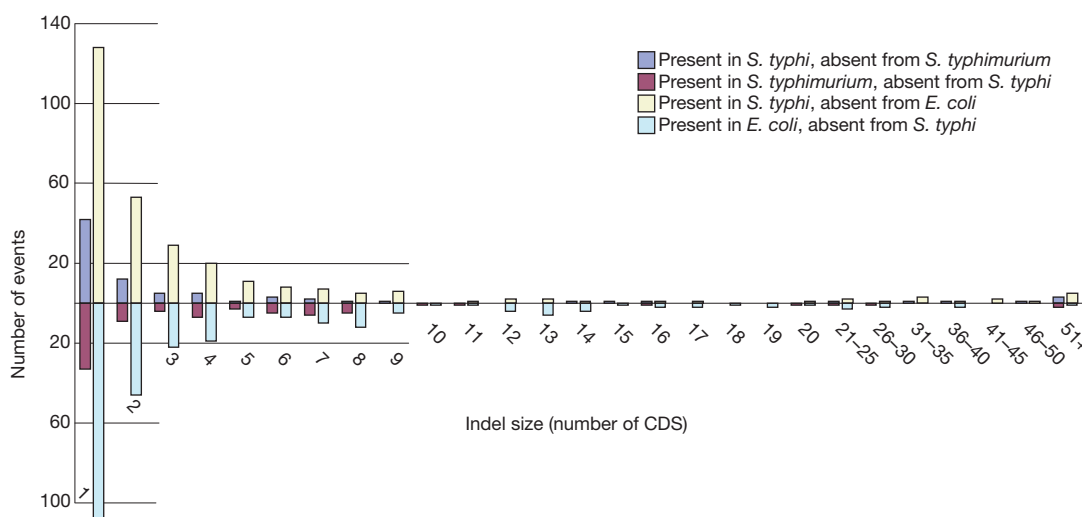


Figure 2 Distribution of insertions and deletions in *S. typhi* relative to *E. coli* and *S. typhimurium*. The graph shows number of insertion/deletion events plotted against the size of the inserted or deleted element (shown as number of genes), clearly indicating that

most of the events involve a small number of genes. Values above the lines represent genes present in *S. typhi*; values below the line represent genes absent in *S. typhi*. Dark bars show the comparison with *S. typhimurium*; light bars with *E. coli*.

and mercury resistance operons have been acquired independently in each. Many clinically relevant resistance genes were identified, including *dhfr1b* (trimethoprim), *sulII* (sulphonamide), *catI* (chloramphenicol), *bla* (TEM-1; ampicillin) and *strAB* (streptomycin). The last four resistance determinants appear to have been inserted into the *tetC* gene of the R27 tetracycline resistance operon in several sequential IS element-mediated events (Fig. 3c). Although it has been reported that MDR plasmids bestow enhanced clinical virulence²³, we can find no obvious virulence-associated genes on pHCM1.

pHCM2 is phenotypically cryptic, yet it shares over 56% of its sequence (at greater than 97% DNA identity) with the *Yersinia pestis* virulence-associated plasmid pMT1 (ref. 24). pMT1 encodes major virulence-associated determinants of *Y. pestis*, and the acquisition of this plasmid was a significant step in the evolution of the plague bacilli²⁵. However, pHCM2 lacks the capsular antigen operon (*cafI*) and murine toxin genes that are characteristic of *Y. pestis* pMT1. The sequences that are shared between pMT1 and pHCM1 are not

contiguous but fall into several blocks (Fig. 3). Examination of the G+C content of the unique and conserved sequences of pMT1 and pHCM2 suggests that pMT1 may have been derived from a pHCM2-like precursor plasmid²⁶. We have detected plasmids related to pHCM2 in *S. typhi* only from Southeast Asia, but most *S. typhi* do not harbour this plasmid (data not shown). The CDS that are unique to pHCM2 show similarities to a number of bacteriophage genes and to genes directly or indirectly involved in DNA biosynthesis and replication. These include a gene cluster that encodes genes similar to thymidylate synthetase, dihydrofolate reductase, ribonuclease H and ribonucleotide diphosphate reductase, as well as a putative primosomal gene cluster. In bacteriophage T4 these genes form an integral part of the primase replication complex²⁷ that facilitates rapid phage DNA biosynthesis and replication.

The sequence of the *S. typhi* genome, together with *E. coli* K12 (ref. 14) and 0157:H7 (ref. 16), reveals an unexpectedly large diversity in gene complements among these organisms. Much of

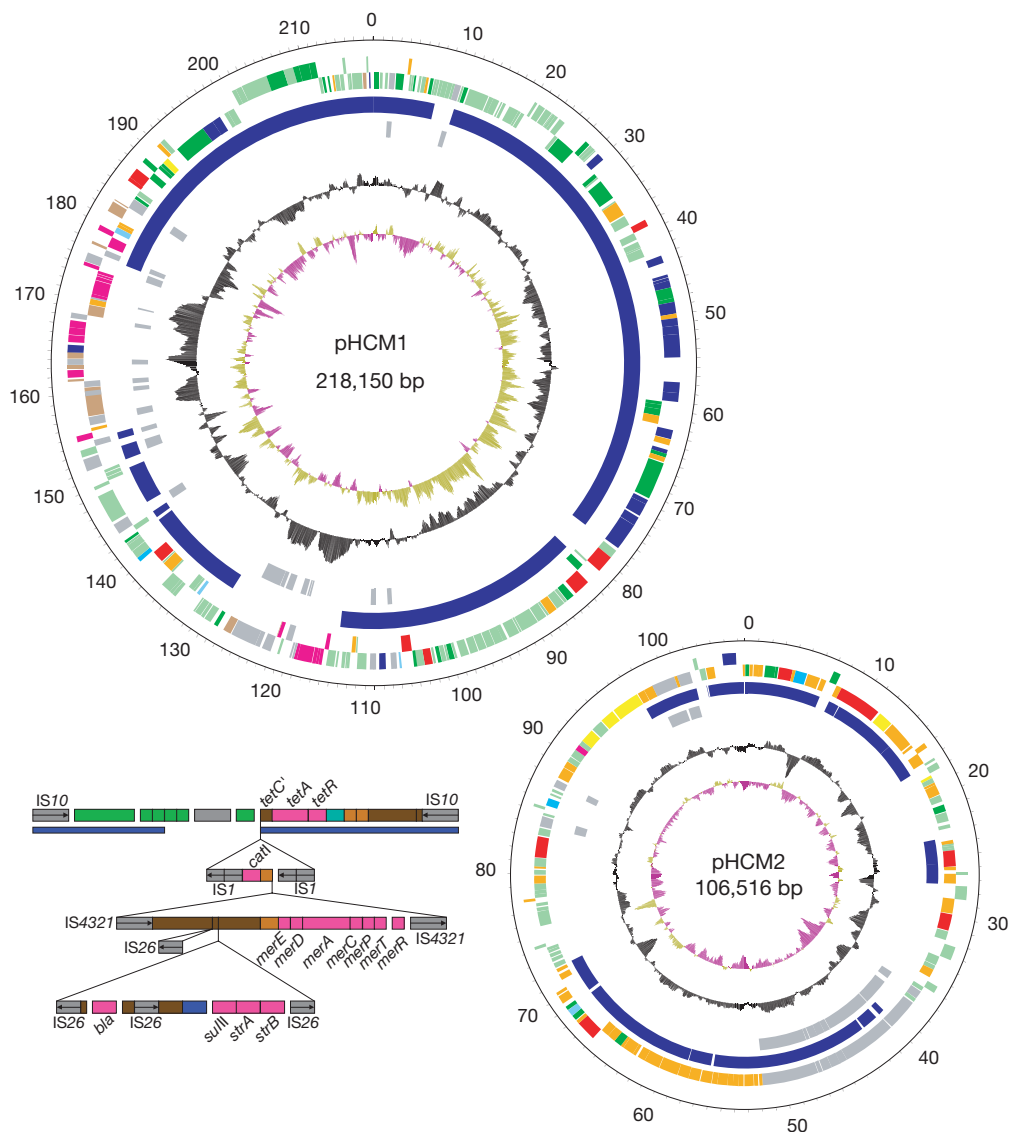


Figure 3 Circular representations of pHCM1 and pHCM2. For the circular diagrams, the outer scale is marked in kilobases. Circles are numbered to the same scheme as in Fig. 1. Circles 1 and 2, all genes; circle 3, regions conserved in R27 (for pHCM1) or pMT1 (for pHCM2); circle 4, phage and IS genes; circle 5, G+C content; circle 6, GC bias ((G – C/G + C); khaki indicates values >1; purple <1). The linear figure represents one possible order of IS-mediated acquisition of resistance determinants in pHCM1 on the

basis of likely pairs of IS elements and interrupted genes. Colour-coding for genes: pale green, unknown function; orange, conserved hypothetical; grey, bacteriophage/IS elements; brown, pseudogenes; red, DNA metabolism; dark green, membrane/surface structures; yellow, metabolic genes; cyan, nucleases; purple, resistance determinants; dark blue, plasmid functions.

this diversity is located in discrete gene clusters that are spread throughout the different genomes. In contrast to this diversity, these enteric microorganisms exhibit marked synteny in their large-scale genomic organization, bearing in mind that *E. coli* and *S. enterica* diverged about 100 Myr ago²⁸. The conserved genes may be a reflection of the basic lifestyle of the bacteria, requiring intestine colonization, environmental survival and transmission. The unique gene clusters probably contribute to adaptation to environmental niches and to pathogenicity. The pseudogene complement of *S. typhi* has implications for our understanding of the tight host restriction of this organism, and raises the question of whether it may be possible to eradicate *S. typhi* and typhoid fever altogether. □

Methods

Salmonella typhi CT18 was isolated in December 1993, at the Mekong Delta region of Vietnam, from a 9-year-old girl who was suffering from typhoid. The strain was isolated from blood using routine culture methods²³, and after serological and metabolic confirmation of the strain as *S. typhi* it was immediately frozen in glycerol at -70 °C. The genome sequence was obtained from 97,000 end sequences (giving 7.9× coverage) derived from several pUC18 genomic shotgun libraries (with insert sizes ranging from 1.4 to 4.0 kb) using dye terminator chemistry on ABI377 automated sequencers. This was supplemented with 0.7× sequence coverage from M13mp18 libraries with similar insert sizes. End sequences from a larger insert plasmid (pSP64; 1.9× clone coverage, 10–14-kb insert size) and lambda (lambda-FIX-II; 0.4× clone coverage, 20–22-kb insert size) libraries were used as a scaffold, and the final assembly was verified by comparison with restriction-enzyme digest patterns using pulsed-field gel electrophoresis (data not shown). Total sequence coverage was 9.1×. The sequence was assembled, finished and annotated as described²⁹, using Artemis³⁰ to collate data and facilitate annotation. In addition we used a genefinder that was trained specifically for *S. typhi*, which uses a hidden Markov model with modules for the coding region, start and stop codons, and the ribosome-binding site (T.S.L. and A.K., unpublished data). The genome and proteome sequences of *S. typhi* and *S. typhimurium* or *E. coli* were compared in parallel to identify deletions and insertions using the Artemis Comparison Tool (ACT) (K. Rutherford, unpublished data; see also <http://www.sanger.ac.uk/Software/ACT/>). Pseudogenes had one or more mutations that would ablate expression, and were identified by direct comparison with *S. typhimurium*; each of the inactivating mutations was subsequently checked against the original sequencing data.

Received 4 April; accepted 20 July 2001.

- Ivanhoff, B. Typhoid fever, global situation and WHO recommendations. *Southeast Asian J. Trop. Med. Public Health* **26**, (Suppl. 2) 1–6 (1995).
- Reeves, M. W., Evans, G. M., Heiba, A. A., Plikaytis, B. D. & Farmer, J. J. Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. *J. Clin. Microbiol.* **27**, 313–230 (1989).
- Parry, C., Wain, J., Chinh, N. T., Vinh, H. & Farrar, J. J. Quinolone-resistant *Salmonella typhi* in Vietnam. *Lancet* **351**, 1289 (1998).
- Neidhardt, F. C. & Curtiss, R. *Escherichia coli* and *Salmonella: Cellular and Molecular Biology* (ASM Press, Washington DC, 1996).
- Rosenberg, S. M., Longrich, S., Gee, P. & Harris, R. S. Adaptive mutation by deletions in small mononucleotide repeats. *Science* **265**, 405–407 (1994).
- Townsend, S. M. *et al.* *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect. Immun.* **69**, 2894–2901 (2001).
- Hensel, M., Hensley, A. P., Nikolaus, T., Sawers, G. & Berks, B. C. The genetic basis of tetrathionate respiration in *Salmonella typhimurium*. *Mol. Microbiol.* **32**, 275–287 (1999).
- Blanc-Potard, A. B., Solomon, F., Kayser, J. & Groisman, E. A. The SPI-3 pathogenicity island of *Salmonella enterica*. *J. Bacteriol.* **181**, 998–1004 (1999).
- Tsolis, R. M. *et al.* Identification of a putative *Salmonella enterica* serotype typhimurium host range factor with homology to IpaH and YopM by signature-tagged mutagenesis. *Infect. Immun.* **67**, 6385–6393 (1999).
- Miao, E. A. & Miller, S. I. A conserved amino acid sequence directing intracellular type III secretion by *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA* **97**, 7539–7544 (2000).
- Bakshi, C. S. *et al.* Identification of SopE2, a *Salmonella* secreted protein which is highly homologous to SopE and involved in bacterial invasion of epithelial cells. *J. Bacteriol.* **182**, 2341–2344 (2000).
- Kingsley, R. A. & Baumber, A. J. Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. *Mol. Microbiol.* **36**, 1006–1014 (2000).
- Sanderson, K. E. & Liu, S. L. Chromosomal rearrangements in enteric bacteria. *Electrophoresis* **19**, 569–572 (1998).
- Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
- Marcus, S. L., Brumell, J. H., Pfeifer, C. G. & Finlay, B. B. *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes Infect.* **2**, 145–156 (2000).
- Perna, N. T. *et al.* Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533 (2001).
- Wong, K. K. *et al.* Identification and sequence analysis of a 27-kilobase chromosomal fragment containing a *Salmonella* pathogenicity island located at 92 minutes on the chromosome map of *Salmonella enterica* serovar typhimurium LT2. *Infect. Immun.* **66**, 3365–3371 (1998).
- Hashimoto, Y., Li, N., Yokoyama, H. & Ezaki, T. Complete nucleotide sequence and molecular characterization of ViB region encoding Vi antigen in *Salmonella typhi*. *J. Bacteriol.* **175**, 4456–4465 (1993).
- Mirol, S. *et al.* Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc. Natl Acad. Sci. USA* **96**, 9845–9850 (1999).

- Zhang, X. L. *et al.* *Salmonella enterica* serovar typhi uses type IVB pili to enter human intestinal epithelial cells. *Infect. Immun.* **68**, 3067–3073 (2000).
- McClelland, M. *et al.* Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, **413**, 852–856 (2001).
- Sherburne, C. K. *et al.* The complete DNA sequence and analysis of R27, a large IncHI plasmid from *Salmonella typhi* that is temperature sensitive for transfer. *Nucleic Acids Res.* **28**, 2177–2186 (2000).
- Wain, J. *et al.* Quantitation of bacteria in blood of typhoid fever patients and relationship between counts and clinical features, transmissibility, and antibiotic resistance. *J. Clin. Microbiol.* **36**, 1683–1687 (1998).
- Hu, P. *et al.* Structural organization of virulence-associated plasmids of *Yersinia pestis*. *J. Bacteriol.* **180**, 5192–5202 (1998).
- Perry, R. D. & Fetherston, J. D. *Yersinia pestis*—etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35–66 (1997).
- Prentice, M. B. *et al.* *Yersinia pestis* pFra shows biovar-specific differences and recent common ancestry with a *Salmonella enterica* serovar Typhi plasmid. *J. Bacteriol.* **183**, 2586–2594 (2001).
- Jing, D. H., Dong, F., Latham, G. J. & von Hippel, P. H. Interactions of bacteriophage T4-coded primase (gp61) with the T4 replication helicase (gp41) and DNA in primosome formation. *J. Biol. Chem.* **274**, 27287–27298 (1999).
- Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G. & Little, E. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**, 470–477 (1996).
- Parkhill, J. *et al.* Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502–506 (2000).
- Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We would like to acknowledge the support of the Sanger Centre core sequencing and informatics groups. We are grateful to the *S. typhimurium* sequencing group at Washington University for sharing sequence and analysis data before publication. A.K. and T.S.L. were supported by a grant from the Danish National Research Foundation. This work was supported by the Wellcome Trust through its Beowulf Genomics initiative and by a Wellcome Trust Programme Grant to G.D.

Correspondence and requests for materials should be addressed to J.P. (e-mail: parkhill@sanger.ac.uk) or G.D. (e-mail: g.dougan@ic.ac.uk). The sequences have been submitted to EMBL with the accession numbers AL513382 (chromosome), AL513383 (pHCM1) and AL513384 (pHCM2).

Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2

Michael McClelland*, Kenneth E. Sanderson†, John Spieth‡, Sandra W. Clifton‡, Phil Latreille‡, Laura Courtney‡, Steffen Porwollik*, Johar Ali‡, Mike Dante‡, Feiyu Du‡, Shunfang Hou‡, Dan Layman‡, Shawn Leonard‡, Christine Nguyen‡, Kelsi Scott‡, Andrea Holmes‡, Neenu Grewal‡, Elizabeth Mulvaney‡, Ellen Ryan‡, Hui Sun‡, Liliana Florea§, Webb Miller§, Tamberlyn Stoneking‡, Michael Nhan‡, Robert Waterston‡ & Richard K. Wilson‡

* Sidney Kimmel Cancer Center, 10835 Altman Row, San Diego, California 92121, USA

† Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

‡ Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St Louis, Missouri 63108, USA

§ Department of Computer Science and Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Salmonella enterica subspecies I, serovar Typhimurium (*S. typhimurium*), is a leading cause of human gastroenteritis, and is used as a mouse model of human typhoid fever¹. The incidence of non-typhoid salmonellosis is increasing worldwide^{2–4}, causing millions of infections and many deaths in the human population each year. Here we sequenced the 4,857-kilobase (kb) chromosome

¶ Present address: Celera Genomics, Rockville, Maryland 20850, USA.

and 94-kb virulence plasmid of *S. typhimurium* strain LT2. The distribution of close homologues of *S. typhimurium* LT2 genes in eight related enterobacteria was determined using previously completed genomes of three related bacteria, sample sequencing of both *S. enterica* serovar Paratyphi A (*S. paratyphi* A) and *Klebsiella pneumoniae*, and hybridization of three unsequenced genomes to a microarray of *S. typhimurium* LT2 genes. Lateral transfer of genes is frequent, with 11% of the *S. typhimurium* LT2 genes missing from *S. enterica* serovar Typhi (*S. typhi*), and 29% missing from *Escherichia coli* K12. The 352 gene homologues of *S. typhimurium* LT2 confined to subspecies I of *S. enterica*—containing most mammalian and bird pathogens³—are useful for studies of epidemiology, host specificity and pathogenesis. Most of these homologues were previously unknown, and 50 may be exported to the periplasm or outer membrane, rendering them accessible as therapeutic or vaccine targets.

The genus *Salmonella* comprises two species: *S. enterica*, which is subdivided into over 2,000 serovars, and *Salmonella bongori*. Some serovars of *S. enterica*, such as *S. typhi*, cause systemic infections and typhoid fever, whereas others, such as *S. typhimurium*, cause gastroenteritis. Some serovars, such as *S. typhi*, are host specialists that infect only humans, whereas others such as *S. typhimurium*, are host generalists that occur in humans and many other mammalian species. Domestic animals act as a reservoir for the food-borne spread of host-generalist serovars, which accounts for the high incidence of non-typhoid *Salmonella* infections worldwide. Estimated costs of food-borne diseases in the United States (with salmonellosis a major component) range from 4.8 to 23 billion dollars³.

Salmonella typhimurium strain LT2, the principal strain for cellular and molecular biology in *Salmonella*, was isolated in the 1940s and used in the first studies on phage-mediated transduction¹. Attenuated mutants of *S. enterica* may be used as live oral vaccines against *Salmonella* infection, to express antigens from other pathogens, and to deliver proteins to solid tumours⁶.

The general characteristics of the *S. typhimurium* LT2 genome are summarized in Table 1. The full dataset is presented as Supplementary Information and is available also at <http://genome.wustl.edu/gsc/Projects/S.typhimurium>.

The overall similarity of *S. typhimurium* LT2 to eight other enterobacterial genomes is summarized in Table 2. We compared *S. typhimurium* LT2 to three other fully sequenced genomes: *S. typhi*⁷, the principal cause of typhoid in humans; *E. coli* K12 (ref. 8), a non-pathogen; and *E. coli* O157:H7, an enterohaemorrhagic strain⁹. (*Escherichia coli* is a member of the closest known genus to *Salmonella*.) We sequenced genome samples over 97% coverage from *S. enterica* serovar *S. paratyphi* A, a frequent cause of typhoid, and *K. pneumoniae*, an opportunistic human pathogen closely related to *Salmonella*. Visual representations of the comparisons of *S. typhimurium* LT2 to the sequenced and sampled genomes are available at both gene and sequence resolution using STM-Enteric and STM-Menteric web tools (<http://galapagos.cse.psu.edu/enterix>)^{10,11}. A total of 4,330 complete open reading frames (ORFs) from the *S. typhimurium* LT2 genome were amplified using specific PCR primers with appended, common 5' ends. These ORFs were microarrayed on glass and probed with fluorescently labelled

Table 2 Coding sequence homologues shared by *S. typhimurium* LT2 and eight other strains of enterobacteria

Organism	Data source	Homologues of STM CDS (%) ^a	Median homology of reciprocal best hit CDS (%)	
			DNA	Amino acid
<i>S. enterica</i> serovar <i>S. typhimurium</i> LT2	Complete sequence	100	100	100
<i>S. typhi</i> CT18	Complete sequence ⁷	89	98	99%
<i>S. paratyphi</i> A	~97% sequence sample/microarray	87/89	98	99
<i>S. paratyphi</i> B	Microarray	92	—	—
<i>S. arizonae</i>	Microarray	83	—	—
<i>S. bongori</i>	Microarray	85	—	—
<i>E. coli</i> K12	Complete sequence ⁸	71	80	90
<i>E. coli</i> O157:H7	Complete sequence ⁹	73	80	89
<i>K. pneumoniae</i>	~97% sequence sample	73	76	88

^aFor sequenced genomes, reciprocal best hits, excluding unsampled regions; for microarrays, signal ratio of *S. typhimurium* LT2 with genome is 3:1 or greater and based on roughly 4,330 *S. typhimurium* LT2 CDS. Raw data is available at <http://genome.wustl.edu/gsc/Projects/S.typhimurium>.

genomic DNA¹² from *S. bongori* and from *S. enterica* serovars *S. paratyphi* A, *S. paratyphi* B and *S. arizonae*, to determine the presence of homologues of *S. typhimurium* LT2 genes. The microarray data for *S. paratyphi* A was over 98% concordant with data from the *S. paratyphi* A genomic sequence (see Supplementary Information). *Salmonella typhimurium* LT2, *S. typhi*, *S. paratyphi* A and *S. paratyphi* B are all in subspecies I of *S. enterica*, which colonizes mammals and birds and causes 99% of *Salmonella* infections in humans^{5,13}. *Salmonella arizonae* is in subspecies IIIa, which colonizes reptiles and also, rarely, humans; *S. bongori*, the only other species of *Salmonella*, does not cause disease in mammals.

Genomic comparisons among the four completed genomes (*S. typhimurium* LT2, *S. typhi*, *E. coli* K12 and *E. coli* O157:H7) reveal that they are collinear for most genes except for inversions over the terminus of replication (TER)^{9,14}. *Salmonella typhimurium* LT2 has a 588-kb inversion compared with *E. coli* K12. Rearrangements between the *rrn* operons are common in *S. typhi*^{7,15}, but are not found in *E. coli* K12, *E. coli* O157:H7 or *S. typhimurium* LT2, or in most other *Salmonella* strains. Large, stable duplications are rare in enterobacterial genomes other than structural RNA genes and transposable elements. In *S. typhimurium* LT2 the largest duplicated region of coding sequences (CDS) is the cytochrome *c* biogenesis locus (*ccm*, 7.5 kb). The two copies are 99% identical at the DNA level and 100% identical for amino acids.

The chromosomes of enteric bacteria are mosaics, composed of collinear regions interspersed with 'loops' or 'islands' unique to certain species; the islands sometimes encode pathogenicity functions (called *Salmonella* pathogenicity islands, SPIs)^{9,16}. Of the 4,489 CDS and pseudogenes annotated in the *S. typhimurium* LT2 chromosome, at least 2,466 (55%) have a close homologue in all eight other enterobacterial genomes that we examined. These homologous genes (shown in red in Fig. 1) along with structural RNAs constitute 2.5 Mb of the *S. typhimurium* LT2 genome.

Table 3 summarizes the distribution of the larger islands among the eight genomes, on the basis of sequence and microarray analysis. SPIs 1–5, previously described in *S. typhimurium* (see refs 1, 17), are listed along with many newly detected large islands and the subset of smaller islands that encode fimbriae or potential virulence factors. Fifteen of the islands are adjacent to a transfer RNA. In fact, almost half of the individually encoded tRNAs are adjacent to an island. At least seven islands encode integrase-like proteins, which are often found in islands¹⁶. *Salmonella typhimurium* LT2 contains four

Table 1 Genome essentials

Parameter	Chromosome	Plasmid pSLT
Size (bp)	4,857,432	93,939
G+C content	53%	53%
rRNA clusters	7	0
tRNAs	85	0
tRNA pseudogene	1	0
Structural RNAs	11	1
CDS (including pseudogenes)	4,489	108
CDS pseudogenes	39	6

functional prophages: Gifsy-1 and -2 (known to have a role in infection^{18,19}) and Fels-1 and -2 (ref. 12). The insertion site of these prophages was predicted from the sequence and was confirmed by microarray-based comparison of labelled DNA from strains that were cured of prophage. These phage are not present in the other eight genomes, although homologues of a few genes exist, presumably in related phage. A previously unknown phage, or phage remnant, that included the *S. typhimurium* LT2 genes *STM4201* to *STM4219* was detected in *S. typhimurium* LT2 by homology to other phage.

Most strains of *S. typhimurium* contain a plasmid of about 90 kb. The plasmid of strain LT2 is called pSLT²⁰. Out of 108 annotated CDS and pseudogenes in pSLT, only three have a close homologue in *S. typhi*, *S. paratyphi* A or *S. paratyphi* B, as expected owing to these strains lacking the plasmid. A search through GenBank revealed 50 pSLT genes that have a close homologue in plasmids from other *Salmonella* serovars. Many homologues of genes in the *tra* operon of the F-factor of *E. coli* K12 were identified, which presumably are responsible for the self-transmissibility of pSLT at rates of up to 3×10^{-4} (ref. 21). The copy number of the pSLT plasmid was estimated as 2.75, using the relative sequence coverage of the plasmid versus the chromosome in the shotgun phase of sequencing, and estimated as 1.4–3.1 under a variety of growth conditions, when measured by average signal intensity on microarrays (data not shown).

Fimbriae on the cell surface mediate adhesion to host cells²². *Salmonella typhimurium* LT2 contains 12 putative operons of the chaperone–usher assembly class: *stc* (called *yehABCD* in *E. coli* K12), *bcf*, *fim*, *lpf*, *saf*, *stb*, *std*, *stf*, *sth*, *sti*, *stj*, all of which are located on the chromosome, and *pef*, which is located on the plasmid. Operons *bcf*, *fim*, *lpf* and *pef* were reported earlier for *S. typhimurium* LT2 and shown to be functional; *saf*, *stb*, *std* and *sth* were detected only by hybridization²². The *sti* and *stj* operons were previously undetected; thus, six of the operons were not previously sequenced and two were not previously detected. *Salmonella typhimurium* LT2 also has the *csg* operon (originally called *agf*) for curli fimbriae, from the nucleator-dependent assembly pathway, but genes for type IV fimbriae⁷ are not detected. Table 3 describes the taxonomic distribution of close homologues of the fimbriae gene clusters in the other eight enterobacterial genomes, but does not include information on the presence of more divergent homologues or orthologues in the other genomes.

Complete sequencing of many closely related genomes, such as *E. coli* K12, *E. coli* O157:H7, *S. typhi* and *S. typhimurium* LT2 (see Table 2), aids the detection of pseudogenes, because a frameshift or stop codon is recognizable only if the gene is collinear with a functional, homologous gene in another genome. This allowed the detection of at least 204 pseudogenes in *S. typhi*⁷, whereas the *S. typhimurium* LT2 chromosome has only about 39. The large number of pseudogenes in *S. typhi* may contribute to or be a consequence of the restriction of *S. typhi* to growth in humans alone⁷, whereas *S. typhimurium* LT2, with a broad host range, has far fewer pseudogenes. Pseudogenes may be unrecognized when a close, intact homologue is unavailable, as is true for 11% or more of the *S. typhimurium* LT2 and *S. typhi* genomes (Table 2). Other potential pseudogenes are encoded across insertion/deletion events that distinguish *S. typhimurium* LT2 and *S. typhi*, such as *S. typhimurium* LT2 gene *STM0098*, which is split by an insertion of about 2,000 bp in *S. typhi* and *S. paratyphi* (or a deletion of this size in *S. typhimurium* LT2). Such genes have been annotated as insertion/deletion events rather than pseudogenes (see Supplementary Information). Finally, pseudogenes might actually have at least one functional fragment. Nevertheless, these caveats do not alter the conclusion that *S. typhimurium* LT2 has far fewer pseudogenes than its close relative *S. typhi*.

The consequences of loss of function in pseudogenes in *S. typhimurium* LT2 is usually unclear, as the function of the known, intact homologue in another organism is often unknown. However, there are pseudogenes in *S. typhimurium* LT2 for maltose

regulation (*malXY*²³) and for trehalose metabolism (*treB*), where the normal allele can substitute in the maltose pathway²⁴; thus, *S. typhimurium* LT2 has disrupted maltose pathways with multiple, independent mutations. Histidine is not used as a carbon source in LT2, although all of the genes are present; this may be explained by a pseudogene mutation in *hutU*. Some of the pseudogenes are in potentially redundant systems: the *dcoA* pseudogene shows over 98% amino-acid identity to another putative sodium ion pump, oxaloacetate decarboxylase alpha chain (*STM3352*); *cutF*

Table 3 Islands and prophages

Start (by STM gene number)	Number of genes	Homologues	Comments
2584*	53	†	Gifsy-1 phage
1005*	52	†	Gifsy-2 phage, <i>sodC</i> , <i>msgA</i> homologue
2865	50	SAL‡	SPI1. Type III secretion system
2694*	47	†	Fels-2 phage
1379§	44	SAL‡	SPI2. Type III secretion system, part missing in SBO
0266*§	42	SS1†	<i>saf</i> fimbriae, <i>virG</i> , <i>sapA</i> , <i>tia</i> and <i>clpB</i> homologues
0893*	40	†	Fels-1 phage, <i>sodC</i> homologue
2019	40	SEN§, KPN	<i>cbi</i> and <i>pdu</i> clusters
0328	37	SS1†	<i>stb</i> fimbriae, <i>ail</i> homologue
1528	35	SAL‡	<i>pqaA</i>
1239§	31	SAL‡	<i>envEF</i> , <i>msgA</i> , <i>pagCD</i>
2741*	27	STM†, KPN‡	Adjacent Fels-2
0014	25	SAL‡	<i>bcf</i> fimbriae
4195	25	†	Many phage homologues
2770	19	SAL‡	<i>fljAB</i> , <i>hin</i> , <i>iroDEN</i> , <i>virK</i> , <i>pipB</i> homologue
4418	19	SPB†	<i>srj</i>
1854*	18	SAL‡	<i>sopE2</i> , <i>pagOK</i>
3117§	18	SAL‡	
3191§	18	SAL	
2230§	16	SPB‡	<i>oafA</i> , <i>sspH2</i> , <i>msgA</i> homologues
4305§	16	SAL‡	<i>phoN</i>
1911	15	SAL, <i>E. coli</i>	<i>flh</i> , flagella, <i>che</i> , <i>mot</i>
1171	14	SAL‡	<i>flg</i> flagella
0715	13	SS1†	Cell wall biogenesis?
3752§	13	SS1†, SBO‡	SPI3 (part) <i>mgtBC</i>
0052	12	SS1†, KPN‡	<i>citDEF</i>
2082	12	SS1	
3025§	12	SS1†	<i>std</i> fimbriae
4440	11	SAL‡, KPN	
0543*§	10	SEN†	<i>ftm</i> fimbriae
1629	10	SAL‡	<i>sseJ</i>
1664	10	SAL‡	<i>invC</i> homologue
3766	10	SS1†, SBO‡	SPI3 (part)
4488*§	9	†	
1087§	8	SAL	SPI5. <i>pip</i> cluster and <i>sopB</i>
3351	8	SEN, KPN	
0426	7	SS1, KPN	<i>phn</i>
0571	7	SAL‡, KPN	
1127§	7	SEN	
1139	7	SAL, <i>E. coli</i>	<i>csg</i> curli fimbriae
3527	7	SPB†	
0195	6	SPA, SPB	<i>stf</i> fimbriae
0854	6	SEN	
1491	6	SAL	
2175	6	SAL, ECH	
2340	6	SAL	
2513	6	SAL‡	<i>shdA</i> , <i>sinIH</i>
3079	6	SPB	
3779	6	SS1†	SPI3 (part)
4010	6	STY, SPA	
4257	6	SAL‡	SPI4
4591	6	SAL‡	<i>sth</i> fimbriae
1328	5	SAL‡	<i>wbbH</i>
2357	5	SEN†	<i>oadABG</i>
3636	5	SPB, SBO	<i>lpf</i> fimbriae
4571	5	†	<i>stj</i> fimbriae
0174	4	SPB	<i>sti</i> fimbriae
1113	4	SEN	<i>scsABCD</i>
1599	4	SAL‡	<i>pdgL</i> , <i>ugtL</i> , <i>sifB</i>
2063	4	SAL	<i>phsABC</i> , <i>sopA</i>
2149	4	STY, SPB	<i>stc</i> fimbriae
2396§	4	SEN, KPN	<i>pgtABCEP</i>

Distribution of *S. typhimurium* LT2 CDS in eight other genomes. Not all islands consisting of five genes or fewer are listed. STM, *S. typhimurium* LT2; SPA, *S. paratyphi* A; SPB, *S. paratyphi* B; STY, *S. typhi*; KPN, *K. pneumoniae*; ECH, *E. coli* O157:H7; SBO, *S. bongori*; SS1 includes STM, SPA, SPB and STY; SEN includes SS1 and *S. arizonae*; and SAL includes SEN and *S. bongori*.

* Encodes integrase.

† Some genes in this cluster have close homologues outside the group.

‡ Some genes in this cluster do not have close homologues in all members of the group.

§ Adjacent to a tRNA.

(STM0241), which encodes the copper homeostasis protein, is a pseudogene, but *cutC* (which remains intact) is sufficient for the same function²⁵.

Genes found only in *Salmonella* (see Table 3 for examples) may have been recruited since the divergence from *Escherichia coli* and *Klebsiella*, or they may have been lost from these genera. There are 1,106 *S. typhimurium* LT2 CDS in this class (marked in green in Fig. 1) that have a close homologue in at least one of the other five *Salmonella* examined. Many *S. typhimurium* LT2 genes associated with pathogenesis are in this class, including invasion genes such as *inv* and *prg*; proteins exported by the type III secretion system, such as SopB, SopD, SopE2, SipA, SipB, SipC and SptP; and some secretory system genes, such as members of *ssa* and *sse*. Most of these genes have homologues in *S. bongori*, indicating that these characteristics are maintained by even the most divergent of *Salmonella*. A subset of 352 *S. typhimurium* LT2 CDS (8%) have close homologues in one or more of the other three subspecies I genomes (*S. typhi*, *S. paratyphi* A and *S. paratyphi* B) but not in *S. arizonae*, *S. bongori*, *E. coli* K12, *E. coli* O157:H7 or *K. pneumoniae*. These genes, most of which were unknown previously, may include genes for specialization of subspecies I to warm-blooded hosts. Among these 352 *S. typhimurium* LT2 CDS are *bigA*, *envF*, *shdA*, *sifAB*, *sinH*, *srff*, *srgAB* (homologues of *ail* and *clpB*), and some genes in the fimbrial clusters, *saf*, *stb*, *stc*, *std*, *stf* and *sti*. Only 70 genes (20%) are named, whereas 49% of all CDS are named, indicating that the group has remained largely unstudied. There are 121 *S. typhimurium* LT2 CDS that have no close homologues in any of the other eight genomes, only a few of which are named, including individual genes from the *pef* and *stj* fimbriae, *spvR*, and a homologue of *mvpA*.

CDS rich in A+T are almost threefold over-represented among genes that have no close homologues outside subspecies I, confirming previous observations made with a smaller set of such genes²⁶. It

is not obvious why CDS confined to only some *Salmonella* should so often be A+T-rich. There is no such bias among G+C-rich genes. If some of these genes are of recent foreign origin then perhaps there is a preferential A+T-rich source for recruited genes.

Attenuated *S. typhimurium* has been used both as a vaccine and for expression of heterologous proteins for vaccines against other organisms²⁷. Proteins that are predicted to be exported to the periplasm or outer membrane, or beyond, are of interest for vaccines and therapy, as they would be exposed for targeting. PSORT²⁸, which predicts cellular location from protein sequence, predicts 251 outer-membrane proteins and 347 periplasmic proteins in *S. typhimurium* LT2, 182 of which are missing from *E. coli* K12, *E. coli* O157:H7 or *K. pneumoniae*, and about 50 of which are confined to subspecies I.

Antigens from all of the predicted surface and secreted CDS can be cloned, expressed on cells, and tested for immunological response, as was done recently for CDS conserved among *Neisseria meningitidis*²⁹. As a first step, we have amplified by PCR almost all the *S. typhimurium* LT2 CDS in a manner suitable for expression, and the distribution of close homologues among eight other enterobacterial genomes has been determined for each gene. □

Methods

Genome sequencing

The complete 4.8-Mb genome of *S. enterica* serovar Typhimurium strain LT2 (American Type Culture Collection, ATCC, number 700720) was sequenced using a combination of two approaches. The first approach was fivefold, whole-genome shotgun sampling, and the second was three- to tenfold sampling of gel-purified restriction fragments ranging in size from 43.3 to 570 kb¹⁴. Sonicated and size-fractionated DNA (approximately 1.5 kb) was cloned into M13 and plasmid vectors. Subclones were sequenced using dye-primer and dye-terminator chemistry on ABI 377 and 3700 sequencing machines. We assembled 92,648 sequence reads, representing ~7.7-fold final coverage, using the PHRAP assembly program (P. Green, unpublished data). The assembled genome sequence is in agreement

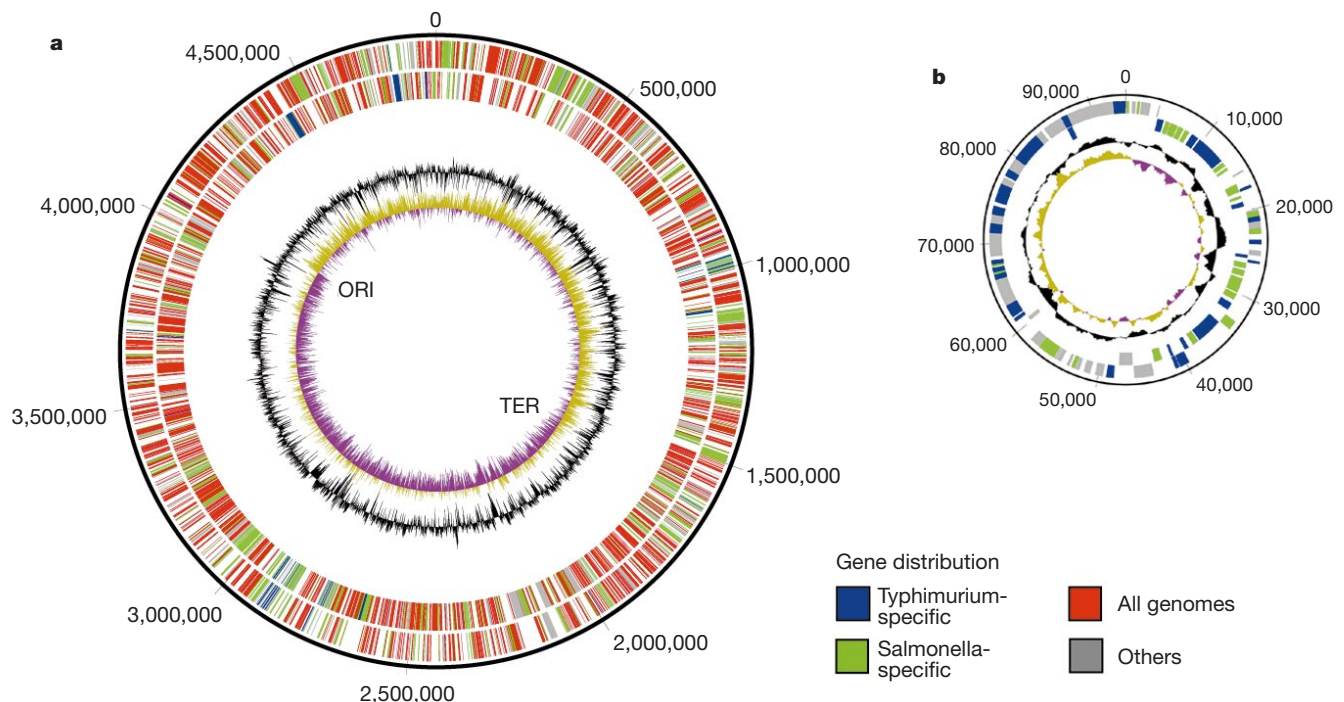


Figure 1 The *Salmonella enterica* serovar Typhimurium LT2 genome. **a**, The chromosome. Base pairs are indicated outside the outer circle. The outer two circles represent the coding orientation, with the forward strand on the outside and the reverse strand on the inside. Red indicates close homologues in all eight genomes. Green indicates genes with a close homologue in at least one other *Salmonella* (*S. typhi*, *S. paratyphi* A, *S. paratyphi* B, *S. arizonae* or *S. bongori*) but not in *E. coli* K12, *E. coli*

O157:H7 and *K. pneumoniae*. Blue indicates genes present only in *S. typhimurium* LT2. Grey indicates other combinations. The black inner circle is the G+C content; the purple/yellow innermost circle is the GC bias. The positions of the origin of replication (ORI) and terminus (TER) are shown. **b**, The plasmid pSLT. Base pairs are indicated outside the outer circle. The plasmid is not to scale. The colouring scheme is the same as for **a**.

with the physical map of the *S. typhimurium* LT2 genome¹⁴, with reads from both ends of about 372 lambda DASH-II clones with inserts of 13–20 kb, and with a *S. typhimurium* LT2 bacterial artificial chromosome library (prepared by R. Wing) restriction finger-printed at the Genome Sequencing Center. Two other genomes were partially sequenced using whole-genome, shotgun sample sequencing: 27,670 sequence reads of *S. enterica* serovar Paratyphi A strain RKS4993 (ATCC 9510) yielded a 3.7-fold coverage (97.5%) for this ~4.6 Mb genome, having 894 contigs with a predicted average gap size of ~200 bp; 36,848 sequence reads of *K. pneumoniae* strain MGH78578 (ATCC 700721), a clinical isolate from sputum, yielded a 3.9-fold depth of coverage (98%) for this ~5.6 Mb genome, resulting in 711 contigs with a predicted average gap size of ~160 bp.

Annotation

The primary annotation database was Acedb (<http://www.acedb.org>). Protein gene predictions were performed using GeneMark (<http://opal.biology.gatech.edu/GeneMark>) and Glimmer (<http://www.tigr.org/softlab/glimmer/glimmer.html>). Predicted proteins were searched against the protein family database, Pfam 5.5 (<http://pfam.wustl.edu>), against the COG database to find putative orthologues in other completed genomes³⁰, and against the protein localization prediction software, PSORT²⁸. The annotation was compared with the *S. typhi* annotation from the Sanger Centre and all differences in CDS predictions and start predictions were reassessed, on the basis of choice of start codon, length, conservation in Enterobacteriaceae and presence of identifiable motifs. The details of microarray construction and hybridization with genomic probes is described elsewhere¹². Strains, lambda and BAC clones are available from the *Salmonella* Genetic Stock Centre, <http://www.ualgary.ca/~kesander>.

Received 10 April; accepted 24 September 2001.

- Neidhardt, F. C. (ed. in chief) *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ASM, Washington DC, 1996).
- Chalker, R. B. & Blaser, M. J. A review of human salmonellosis: III. Magnitude of *Salmonella* infection in the United States. *Rev. Infect. Dis.* **10**, 111–124 (1988).
- Todd, E. Epidemiology of foodborne illness: North America. *Lancet* **336**, 788–790 (1990).
- Cooke, E. M. Epidemiology of foodborne illness: UK. *Lancet* **336**, 790–793 (1990).
- Selander, R. K., Li, J. & Nelson, K. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (eds Neidhardt, F. C. et al.) 2691–2707 (ASM, Washington DC, 1996).
- Bermudes, D., Low, B. & Pawelek, J. Tumor-targeted *Salmonella*. Highly selective delivery vectors. *Adv. Exp. Med. Biol.* **465**, 57–63 (2000).
- Parkhill, J. et al. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848–852 (2001).
- Blattner, F. R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
- Perna, N. T. et al. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533 (2001).
- Florea, L. et al. Web-based visualization tools for bacterial genome alignments. *Nucleic Acids Res.* **28**, 3486–3496 (2000).
- McClelland, M. et al. Comparison of the *Escherichia coli* K-12 genome with sampled genomes of a *Klebsiella pneumoniae* and three *Salmonella enterica* serovars, Typhimurium, Typhi and Paratyphi. *Nucleic Acids Res.* **28**, 4974–4986 (2000).
- Porwollik, S. et al. The Δ -uvrB mutations in the Ames strains of *Salmonella* span 15 to 119 genes. *Mut. Res.* **483**, 1–11 (2001).
- Popoff, M. Y., Bockemuhl, J. & Brenner, F. W. Supplement 1999 (no. 43) to the Kauffmann-White scheme. *Res. Microbiol.* **151**, 893–896 (2000).
- Liu, S. L., Hessel, A. & Sanderson, K. E. The *XbaI*-*BlnI*-*CeuI* genomic cleavage map of *Salmonella typhimurium* LT2 determined by double digestion, end labelling, and pulsed-field gel electrophoresis. *J. Bacteriol.* **175**, 4104–4120 (1993).
- Liu, S. L. & Sanderson, K. E. Highly plastic chromosomal organization in *Salmonella typhi*. *Proc. Natl Acad. Sci. USA* **93**, 10303–10308 (1996).
- Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
- Blanc-Potard, A. B., Solomon, F., Kayser, J. & Groisman, E. A. The SPI-3 pathogenicity island of *Salmonella enterica*. *J. Bacteriol.* **181**, 998–1004 (1999).
- Stanley, T. L., Ellermeier, C. D. & Schlauch, J. M. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar Typhimurium survival in Peyer's patches. *J. Bacteriol.* **182**, 4406–4413 (2000).
- Figuerola-Bossi, N., Uzzau, S., Maloriel, D. & Bossi, L. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol. Microbiol.* **39**, 260–272 (2001).
- Matsui, H. et al. Virulence plasmid borne *spvB* and *spvC* genes can replace the 90-kilobase plasmid in conferring virulence to *Salmonella enterica* serovar Typhimurium in subcutaneously inoculated mice. *J. Bacteriol.* **183**, 4652–4658 (2001).
- Ahmer, B. M., Tran, M. & Heffron, F. The virulence plasmid of *Salmonella typhimurium* is self-transmissible. *J. Bacteriol.* **181**, 1364–1368 (1999).
- Townsend, S. M. et al. *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect. Immun.* **69**, 2894–2901 (2001).
- Reidl, J. & Boos, W. The *malX* *malY* operon of *Escherichia coli* encodes a novel enzyme II of the phosphotransferase system recognizing glucose and maltose and an enzyme abolishing the endogenous induction of the maltose system. *J. Bacteriol.* **173**, 4862–4876 (1991).
- Decker, K., Gerhardt, F. & Boos, W. The role of the trehalose system in regulating the maltose regulon of *Escherichia coli*. *Mol. Microbiol.* **32**, 777–788 (1999).
- Gupta, S. D., Lee, B. T., Camakaris, J. & Wu, H. C. Identification of *cutC* and *cutF* (*nlpE*) genes involved in copper tolerance in *Escherichia coli*. *J. Bacteriol.* **177**, 4207–4215 (1995).
- Lan, R. & Reeves, P. R. Gene transfer is a major factor in bacterial evolution. *Mol. Biol. Evol.* **13**, 47–55 (1996).
- Bumann, D., Hueck, C., Aebischer, T. & Meyer, T. F. Recombinant live *Salmonella* spp. for human vaccination against heterologous pathogens. *FEMS Immunol. Med. Microbiol.* **27**, 357–364 (2000).

- Nakai, K. & Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24**, 34–36 (1999).
- Pizza, M. et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**, 1816–1820 (2000).
- Tatusov, R. L. et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**, 22–28 (2001).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by National Institutes of Health grants to R.W., M.M. and W.M. Nine ORFs were added to the annotation courtesy of L. Umayam and O. White of the Comprehensive Microbial Resource. We thank T. Lynch, M. Becker, G. Duckels and P. Minx for library production; J. Welsh, E. Koonin, R. Tatusov, H. Salgado, J. Collado-Vides and K. Nakai for assistance with computer analysis; S. Sims for assistance with primer design; M. Sekhon and the GSC Mapping Group for BAC mapping; S.-L. Liu for large restriction fragments of *S. typhimurium* LT2; A. McKay, B. Pearson, J. Roth and M. Riley for advice; and J. Parkhill and G. Dougan for sharing data from the *S. typhi* project.

Correspondence and requests for materials should be addressed to M.M. (e-mail: mmccllland@skcc.org). The sequence of the *S. typhimurium* LT2 chromosome and the annotated pSTL plasmid are deposited in GenBank under accession numbers AE006468 and AE006471, respectively.

Frizzled-7 signalling controls tissue separation during *Xenopus* gastrulation

Rudolf Winklbauer*†‡, Araceli Medina*†, Rajeeb K. Swain† & Herbert Steinbeisser*

† Max-Planck-Institute for Developmental Biology, Department of Cell Biology, Spemann Strasse 35, D-72076 Tübingen, Germany

* These authors contributed equally to this work

Cell signalling through Frizzled receptors has evolved to considerable complexity within the metazoans. The Frizzled-dependent signalling cascade comprises several branches, whose differential activation depends on specific Wnt ligands, Frizzled receptor isoforms and the cellular context. In *Xenopus laevis* embryos, the canonical β -catenin pathway contributes to the establishment of the dorsal–ventral axis¹. A different branch, referred to as the planar cell polarity pathway, is essential for cell polarization during elongation of the axial mesoderm by convergent extension². Here we demonstrate that a third branch of the cascade is independent of Dishevelled function and involves signalling through trimeric G proteins and protein kinase C (PKC). During gastrulation, Frizzled-7 (Fz7)-dependent PKC signalling controls cell-sorting behaviour in the mesoderm. Loss of zygotic Fz7 function results in the inability of involuted anterior mesoderm to separate from the ectoderm, which leads to severe gastrulation defects. This result provides a developmentally relevant *in vivo* function for the Fz/PKC pathway in vertebrates.

In *Xenopus*, maternal Fz7 acts in the Wnt/ β -catenin pathway and contributes to dorsalization of the embryo³. To study the role of zygotic *Xenopus* Fz7, we attempted a 'loss-of-function' approach that depends on blocking translation of the Fz7 messenger RNA by an antisense morpholino oligonucleotide directed against the 5'

‡ Present address: Department of Zoology, University of Toronto, 25 Harbord Street, Toronto, Ontario M5S 3G5, Canada.

untranslated region (UTR). First, we demonstrated the efficiency of inhibition (Fig. 1). Translation of Fz7-Myc RNA containing the target site of the antisense morpholino (Fz7Mo) was blocked in a reticulocyte lysate (Fig. 1a). Translation in the embryo was also blocked, as shown by anti-Myc antibody staining of western blots and of animal cap cells (Fig. 1b–e). In contrast, translation of Fz7-Myc RNA that lacked the Fz7Mo target site or of unrelated enhanced green fluorescent protein (EGFP)-Myc RNA was not affected. An unrelated control morpholino oligonucleotide (CoMo) had no inhibitory effects.

Dorsal injection of Fz7Mo, but not of CoMo, at the eight-cell stage caused severe defects (Fig. 1f–i). Embryos at the tailbud stage typically showed spina bifida and a reduction of head development.

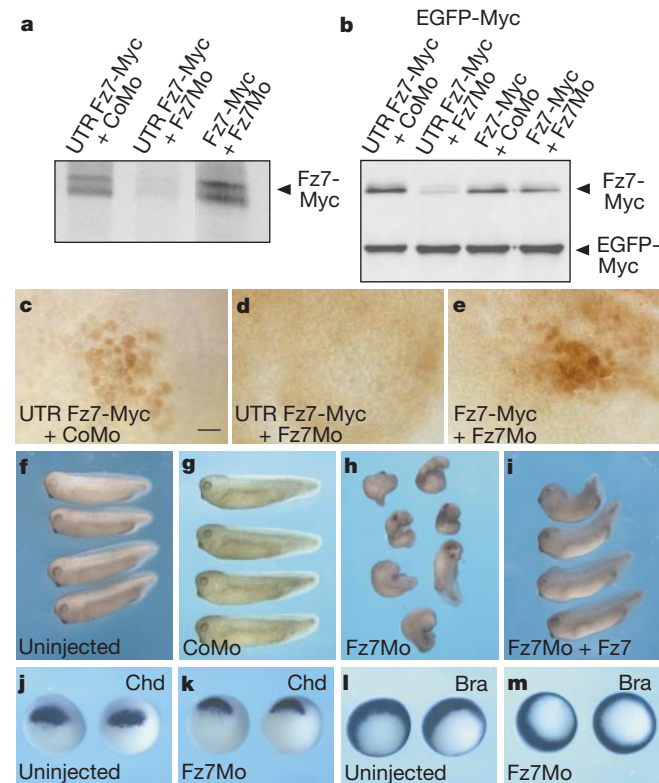


Figure 1 Antisense morpholino oligonucleotide blocks translation of *Xenopus* Fz7 mRNA and impairs morphogenesis. **a**, mRNAs and protein were prepared from 250 ng UTR Fz7-Myc and Fz7-Myc plasmid using an *in vitro* transcription/translation system. ³⁵S-labelled translation products were analysed on SDS-PAGE gels. Translation of UTR Fz7-Myc RNA was blocked in the presence of antisense Fz7 morpholino oligonucleotide (Fz7Mo) but not by control oligonucleotide (CoMo). Translation of Fz7-Myc RNA lacking the 5' UTR sequences could not be blocked by Fz7Mo. **b**, Western blot analysis of Fz7 protein tagged with a Myc epitope at the C-terminus using anti-Myc antibody. Translation of injected EGFP-Myc (5 pg) and Fz7-Myc RNA lacking UTR sequences (300 pg) was not inhibited by either CoMo or Fz7Mo (5 pmol, 40 ng), indicating that the morpholino oligonucleotides do not unspecifically inhibit translation. **c–e**, Similar results were obtained after injection of 200 pg UTR Fz7-Myc or Fz7-Myc expression plasmid DNA in combination with Fz7Mo or CoMo (10 pmol, 80 ng) into the animal region of embryos at the two-cell stage. Fz7-Myc protein was detected at gastrula stages when UTR Fz7-Myc and CoMo were injected (**c**) but not in the combination UTR Fz7-Myc and Fz7Mo (**d**). Fz7Mo was unable to repress translation of Fz7-Myc RNA (**e**). Scale bar, 50 μ m. **f–i**, The four dorsal blastomeres of embryos at the eight-cell stage were injected with 200 pg Fz7 RNA² and 160 ng Fz7Mo or CoMo oligonucleotide. Phenotypes were analysed at stage 35. Uninjected embryos (**f**) and embryos injected with CoMo (**g**) at the tailbud stage are shown. **h**, Embryos injected with Fz7Mo display trunk defects (57/70). **i**, Defects caused by Fz7Mo are partially rescued by Fz7 RNA (200 pg) (33/66). **j–m**, Whole-mount *in situ* hybridization for *chordin* (Chd) (**j**, **k**) and *brachyury* (Bra) (**l**, **m**) at early gastrula stages in uninjected embryos (**j**, **l**) and embryos injected with Fz7Mo (20 pmol, 160 ng) (**k**, **m**).

These effects could be rescued by coinjection of *Xenopus* Fz7 mRNA that lacked the morpholino target site. Expression patterns of mesodermal (*brachyury*) and organizer markers (*chordin*) were not changed in gastrulae that had been injected with Fz7Mo (Fig. 1j–m). This suggests that the phenotypic effects were not due to perturbed mesoderm induction or patterning; instead, morphogenetic processes during gastrulation may be affected.

Convergent extension of the axial mesoderm, a principal gastrulation process, is controlled by Wnt11 acting in the planar cell polarity (PCP) pathway in both *Xenopus* and *Zebrafish*^{4,5}. As *Xenopus* Wnt11 and Fz7 can interact⁶, we examined whether loss of zygotic Fz7 would affect convergent extension (Fig. 2). Explants of the dorsal blastopore lip from uninjected controls, from embryos injected with CoMo or Fz7Mo, and from embryos coinjected with Fz7Mo and Fz7 mRNA, all elongated similarly (Fig. 2b–e). Thus, in contrast to Fz7 overexpression⁷, loss of Fz7 function did not impair convergent extension, indicating that Fz7 is not essentially involved in mediating this aspect of Wnt11 function.

Evidence for a role of *Xenopus* Fz7 in early morphogenesis was obtained from the analysis of gastrulae that were fractured sagittally (Fig. 2f–i), where anterior mesoderm and endoderm had become apposed to the blastocoel roof (BCR) (Fig. 2j). To allow for its further movement, the involuted mesoderm must remain separated from the BCR by Brachet's cleft, despite the absence of an insulating physical barrier and despite the expression of the same cadherin isoforms in both of these tissues⁸. Notably, in embryos that were injected with Fz7Mo, the height of the dorsal blastopore lip was significantly greater than in uninjected or CoMo-injected specimens. Brachet's cleft was only present anteriorly, where vegetal cells about the BCR⁹. The boundary between BCR and involuted mesoderm had disappeared posteriorly (Fig. 2). This suggests a lack of tissue separation in this region, due to the absence of zygotic Fz7 function. The fusion of BCR and involuted anterior mesoderm was prevented by coinjection of Fz7Mo-resistant Fz7 mRNA, demonstrating that the effect was specific.

Involuted mesoderm cells display separation behaviour, which allows them to remain on the surface of the ectodermal BCR, and thus provides the cellular basis for tissue separation⁸. Cells that do not show this behaviour sink into the BCR and mix with the cells there. These different behaviours can be analysed in an *in vitro* assay (Fig. 2k–o). Cell aggregates from the lower part of the dorsal blastopore lip, which had acquired separation behaviour⁸, remained on the surface of explanted BCR. When injected with Fz7Mo, the same cells integrated into the untreated BCR layer. This loss of separation behaviour of lower lip tissue, which was not obtained with CoMo, could be rescued by full-length Fz7, but not by a secreted extracellular domain of the protein (NFz7-fun), and only partially by Fz8. Expression of NFz7-fun alone in lower lip tissue reduced separation behaviour, presumably by competing with endogenous Fz7 for a ligand (Fig. 2o). Thus signalling through Fz7 is required during gastrulation in the anterior mesoderm—but not in the BCR (not shown)—for tissue separation at Brachet's cleft.

Depending on the cellular context, Fz7 is able to activate the β -catenin, the PCP and the Fz/PKC pathways^{3,6,7}. To determine which branch is involved in control of separation behaviour, activating or inhibiting components of the different pathways were ectopically expressed and effects on separation behaviour were analysed in the *in vitro* separation assay (Fig. 2o). A dominant-negative mutant of GSK3, activated transcription factor TCF3 or stabilized β -catenin, which all activate the canonical pathway on overexpression^{10–12}, were unable to restore separation behaviour in mesoderm that was injected with Fz7Mo. This indicates that loss of Fz7 function cannot be compensated for by downstream activation of β -catenin signalling. Similarly, activating components of the PCP branch, such as a Dsh mutant lacking the Dix domain, or an activated form of Cdc42 (Cdc42G12V)⁶, could not rescue separation behaviour in embryos

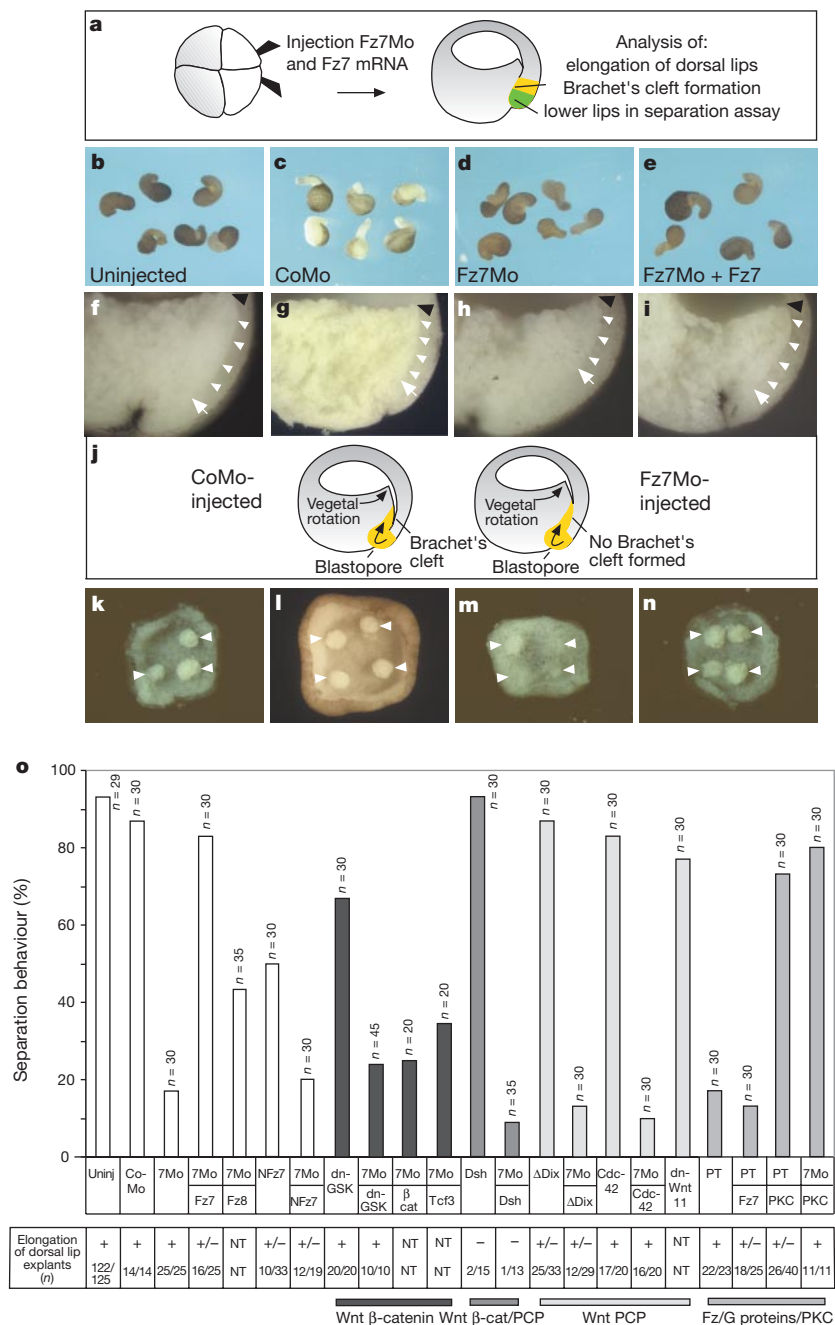


Figure 2 Antisense morpholino oligonucleotide for *Xenopus* Fz7 inhibits tissue separation behaviour but not convergent extension movements. **a**, Embryos (animal view) were injected with CoMo or Fz7Mo (20 pmol, 160 ng) alone or in combination with Fz7 RNA into the four dorsal blastomeres at the eight-cell stage. **b–e**, At stage 10, dorsal lips were explanted and elongation was scored when the embryos had reached late neurula stage 19. Dorsal lip explants from uninjected embryos (**b**), embryos injected with CoMo (**c**), Fz7Mo oligonucleotide (**d**), or with Fz7Mo and Fz7 RNA (**e**) are elongated. **f–i**, Formation of Brachet's cleft on the dorsal side was analysed in embryos at stage 10.5 that were fractured sagittally. The anterior end of Brachet's cleft is indicated by black arrowheads, posterior end by a white arrow, and the extent of the cleft by white arrowheads. Uninjected embryos (**f**), embryos injected with CoMo (**g**), Fz7Mo (**h**), or with Fz7Mo and Fz7 RNA (**i**) are shown. The height of the lip (measured from the blastopore to the posterior end of Brachet's cleft; white arrow) is increased in **h, j**. Summary of the effect of Fz7Mo on the formation of Brachet's cleft. **k–n**, The ability of lower lip mesoderm to separate from BCR tissue was scored in an *in vitro* assay. Explants from uninjected embryos (**k**) and embryos injected with CoMo (**l**) displayed separation behaviour and

stayed on the surface of the BCR, but those from embryos injected with Fz7Mo moved into the BCR layer (**m**). **n**, Separation behaviour could be restored in explants injected with Fz7Mo by Fz7 RNA. White arrowheads indicate the position of explants on the BCR. **o**, The effects of components acting in the Wnt/ β -catenin, PCP and Fz/PCK pathways on separation behaviour of lower lip mesoderm was quantitatively analysed by the assay shown in **k–n**, and elongation of dorsal lip explants was scored. CoMo, control Morpholino oligo (20 pmol, 160 ng); 7 Mo, anti-Xfz7 Morpholino oligo (20 pmol, 160 ng); fz7, *Xenopus* frizzled 7 (RNA, 300 pg); fz8, *Xenopus* frizzled 8 (ref. 19) (RNA, 300 pg); Nfz7, secreted N terminus of Xfz7 fused to the fun-domain of Frzb-1 (ref. 7) (RNA, 300 pg); dnGSK, dominant negative GSK3 mutant¹⁰ (RNA, 1 ng); β -cat, stable β -catenin¹² (RNA, 400 pg); TCF3, activated TCF3 (ref. 11) (RNA, 400 pg); dsh, *Xenopus* dishevelled¹³ (RNA, 500 pg); Δ Dix, dishevelled mutant lacking the Dix domain²⁰ (RNA, 500 pg); Cdc42, activated form of Cdc42 (ref. 6) (DNA, 400 pg); dnWnt11, dominant negative *Xenopus* wnt11 (ref. 4) (RNA, 500 pg); PT, pertussis toxin¹⁴ (RNA, 500 pg); PKC, protein kinase C α ¹⁵ (RNA, 1 ng); nt, not tested. +, elongated explants, –, no elongation, +/-, partial elongation.

that were treated with Fz7Mo. Expression of dominant-negative Wnt11 strongly inhibited convergent extension⁴ but did not suppress separation behaviour. This suggests that the PCP branch of Fz signalling is also not implicated in the control of tissue separation. Dishevelled (Dsh) acts downstream of the Frizzled receptor in both the Wnt/ β -catenin and the PCP pathways^{2,13}. As expected, overexpression of Dsh inhibited convergent extension by interfering with the PCP pathway. Separation behaviour was not affected and was not rescued by Dsh in embryos that were injected with Fz7Mo. Taken together, this suggests that separation behaviour is regulated neither through the canonical nor through the PCP pathway, and in a manner that is independent of Dsh.

The Fz/PKC pathway involves activation of heterotrimeric G proteins of the G_i class^{14,15}. When G-protein-mediated signalling was blocked in the lower lip by expression of pertussis toxin, separation behaviour was strongly repressed. It could not be restored by coexpression of Fz7, but it was rescued by PKC α ¹⁶. Separation behaviour was also rescued by PKC α in lower lip mesoderm that had been injected with Fz7Mo. This suggests that Fz7 acts upstream, and PKC α downstream, of the pertussis-toxin-sensitive G proteins. Consistent with this finding, ectopically expressed PKC α -Myc was activated on the dorsal side in a manner that was dependent on Fz7 (Fig. 3). During activation, PKC α becomes phosphorylated and phosphorylated PKC α protein was detected after injection of PKC α -Myc together with CoMo, but not in combination with Fz7Mo. Coinjection of Fz7 mRNA was able to restore PKC α activation (Fig. 3a–c).

Expression of Fz7 was sufficient to induce separation behaviour ectopically in fibroblast growth factor (FGF)-induced animal cap cells (Fig. 4a, b). In contrast to the small and complex blastopore lip,

this system allowed a convenient further characterization of Fz7-controlled separation behaviour in a homogeneous cell population. As shown previously⁸, aggregates of animal cap cells that were placed on BCR explants integrated rapidly into the substratum layer, and application of FGF did not change this behaviour. Expression of Fz7 alone in animal cap cells induced separation behaviour weakly, but induction was strong in the presence of FGF

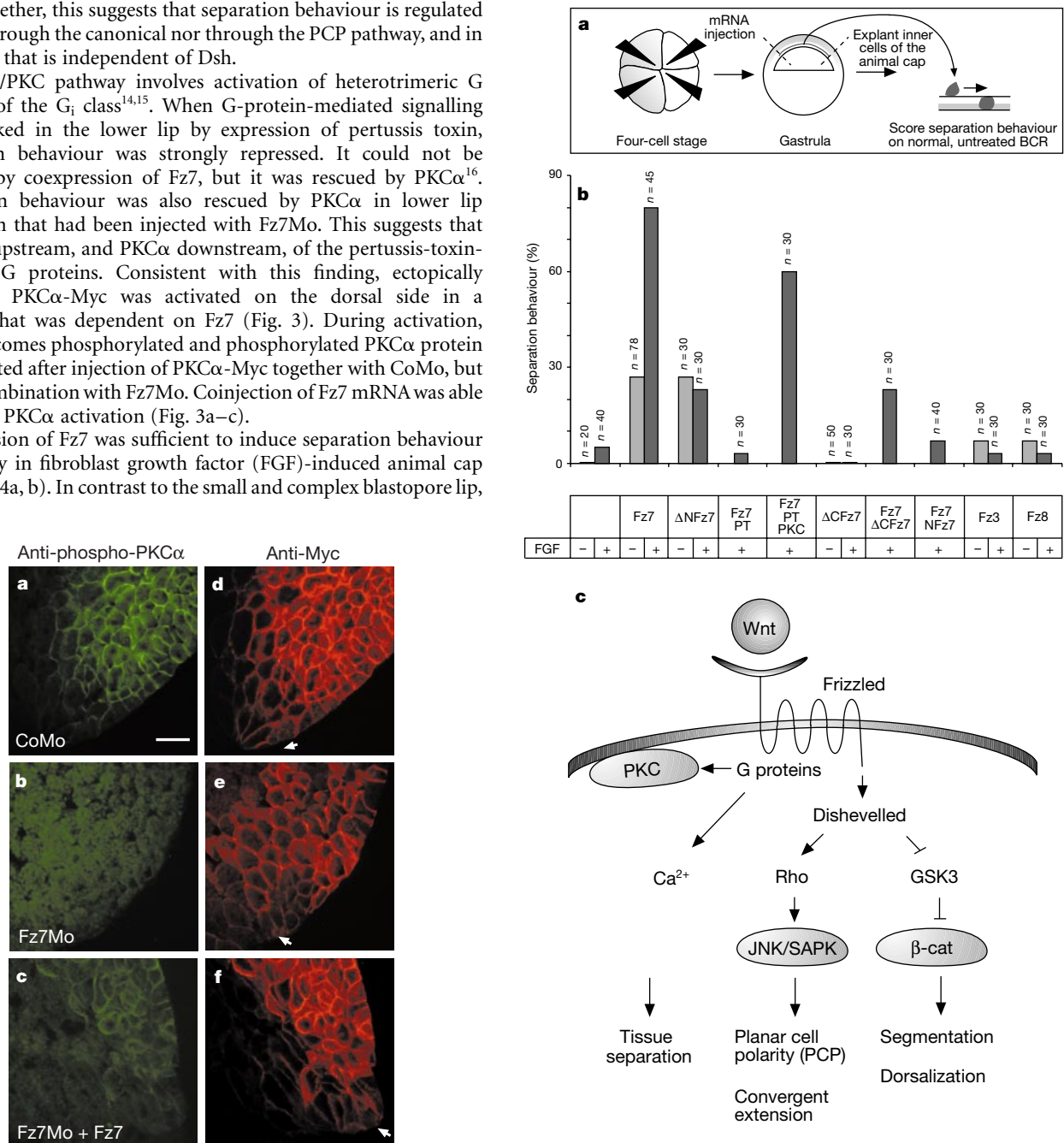


Figure 3 Antisense morpholino oligonucleotide for *Xenopus* Fz7 inhibits phosphorylation of PKC α -Myc. Embryos were injected on the dorsal side at the eight-cell stage with RNAs for PKC α -Myc¹⁵ (400 pg), Fz7 (ref. 7) (300 pg), CoMo (20 pmol, 160 ng) and anti-Fz7 morpholino (20 pmol, 160 ng). **a–f**, Phosphorylated PKC α -Myc was detected using an anti-phospho-PKC α (Ser 657) antibody (**a–c**) and the Myc epitope with the anti-Myc antibody 9E10 (**d–f**) on 10 μ m cryosections. Embryos injected with PKC α -Myc and CoMo (**a, d**), PKC α -Myc and anti-Fz7 morpholino (**b, e**) and PKC α -Myc, anti-Fz7 morpholino oligonucleotide and Fz7 (**c, f**) are shown. Arrowheads indicate dorsal blastopore lip. Scale bar, 50 μ m.

Figure 4 *Xenopus* Fz7 induces separation behaviour in animal cap cells. **a**, Embryos were injected at the four-cell stage into the animal region. At stage 10+, inner animal cap cells that were derived from injected blastomeres were tested on untreated BCRs in the *in vitro* separation assay⁸. **b**, Compilation of the experiments. RNAs for the following proteins were injected: FGF, embryonic fibroblast growth factor¹⁷ (10 pg); Fz7 (ref. 7) (500 pg); Δ NFz7, Fz7 mutant lacking the extracellular N terminus (500 pg); Δ CFz7, Fz7 mutant lacking the intracellular C terminus⁷ (500 pg); NFz7 (500 pg); PT, pertussis toxin¹⁴ (500 pg); PKC α ¹⁵ (1 ng); Fz3 (ref. 21) (500 pg); Fz8 (ref. 19) (500 pg). **c**, Frizzled/PKC signalling regulates tissue separation behaviour in a Dsh-independent manner. JNK, Jun N-terminal kinase. SAPK, stress-activated protein kinase.

signalling. Separation behaviour induced by Fz7 in FGF-stimulated animal cap cells was inhibited by pertussis toxin, and this block was released by coexpression of PKC α . This indicates that, as in the lower lip, the PKC branch of the Fz signalling cascade is involved, and that the animal cap system is a valid model for endogenous involuted mesoderm. Induction of separation behaviour is specific to *Xenopus* Fz7, as similar ectopic expression of Fz3 or Fz8 did not promote it.

The response to Fz7 expression is ligand-dependent. A Fz7 deletion construct lacking the extracellular amino terminus (Δ NFz7) induced a low level of separation behaviour, which could not be further increased by FGF. Coexpression of Fz7 with a secreted, putative ligand-binding domain (NXfz7-fun) or a construct lacking the intracellular carboxy-terminus (Δ CFz7) prevented the separation response in FGF-stimulated animal cap cells, demonstrating that these constructs could act as dominant-negative proteins.

Together, our results with dorsal lips and animal caps show that the PCP and the Fz/PKC branches are clearly separable pathways. In particular, the Fz/PKC pathway does not involve Dsh function (Fig. 4c). During gastrulation this pathway controls tissue separation in the anterior mesoderm through the *Xenopus* Fz7 receptor, which is critical for normal development. This type of boundary formation is an important process in embryogenesis and its regulation is of general significance. Further experiments will clarify whether Fz/PKC signalling is also involved in other tissue separation events in *Xenopus* as well as in other vertebrate and invertebrate species. □

Methods

DNA constructs

Two hundred and one base pairs of the 5' UTR and the complete coding sequence of *Xenopus* Fz7 were amplified by polymerase chain reaction (PCR) and cloned into *Bam*HI/*Cl*AI sites of the expression plasmid pCS2+—MT to generate Myc-tagged UTR Fz7. For the N-terminal deletion of Fz7 (Δ NFz7), a fragment from the coding region (residues 215–549) containing the seven transmembrane domains and the carboxyl tail was amplified by PCR and cloned into the *Eco*RI/*Xba*I sites of the pCS2+ vector. A fragment coding for the signal peptide (residues 1–52) was joined with this clone (residues 215–549) at the *Bam*HI/*Eco*RI sites. Sense RNA for Δ NFz7 was transcribed from template that was linearized with *Asp* 718, using the SP6 promoter.

Sequence of Fz7Mo and transcription/translation of Fz7

Fz7Mo oligonucleotide (5'-CCAACAAGTGATCTCTGGACAG CAG-3') and unspecific control morpholino oligonucleotide (Gene-Tools). The target sequence for Fz7Mo is not present in the 5' UTR and the coding region of the Fz8 complementary DNA. For *in vitro* transcription/translation of Fz7 we used the TNT Coupled Reticulocyte Lysate System (Promega) according to the manufacturer's instructions.

Whole-mount analysis and immunofluorescence

Expression of *brachyury* and *chordin* was analysed in embryos at the gastrula stage as described⁷. Immunostaining and colorimetric detection using horseradish peroxidase has been described¹⁷. Protein extracts were separated on SDS–polyacrylamide gel electrophoresis (PAGE) gels, and EGFP-Myc and Fz7-Myc protein was detected with the 9E10 antibody as described¹⁷.

Embryos were fixed in Dent's solution (methanol/DMSO; 4:1). Cryosections (10 μ m) and immunodetection of the phosphorylated form of PKC α (anti-phospho-PKC α Ser 657; Upstate Biotechnology) and Myc-epitope (9E10) was performed as described¹⁸.

Separation behaviour and dorsal lip elongation

In vitro separation assay was performed as described⁸. Convergent extension in dorsal lips was assayed as described².

Received 26 June; accepted 5 September 2001.

1. Miller, J. R., Hocking, A. M., Brown, J. D. & Moon, R. T. Mechanism and function of signal transduction by the Wnt/ β -catenin and Wnt/Ca²⁺ pathways. *Oncogene* **18**, 7860–7872 (1999).
2. Wallingford, J. B. *et al.* Dishevelled controls cell polarity during *Xenopus* gastrulation. *Nature* **405**, 81–85 (2000).
3. Sumana, S., Strege, P., Heasman, J. & Ekker, S. C. The putative wnt receptor *Xenopus* frizzled-7 functions upstream of β -catenin in vertebrate dorsoventral mesoderm patterning. *Development* **127**, 1981–1990 (2000).
4. Tada, M. & Smith, J. C. Xwnt11 is a target of *Xenopus* Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. *Development* **127**, 2227–2238 (2000).

5. Heisenberg, C. P. *et al.* Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76–81 (2000).
6. Djiane, A., Riou, J., Umbhauer, M., Boucaut, J. & Shi, D. Role of frizzled 7 in the regulation of convergent extension movements during gastrulation in *Xenopus laevis*. *Development* **127**, 3091–3100 (2000).
7. Medina, A., Reintsch, W. & Steinbeisser, H. *Xenopus* frizzled 7 can act in canonical and non-canonical Wnt signaling pathways: implications on early patterning and morphogenesis. *Mech. Dev.* **92**, 227–237 (2000).
8. Wacker, S., Grimm, K., Joos, T. & Winklbauer, R. Development and control of tissue separation at gastrulation in *Xenopus*. *Dev. Biol.* **224**, 428–439 (2000).
9. Winklbauer, R. & Schürfeld, M. Vegetal rotation, a new gastrulation movement involved in the internalization of the mesoderm and endoderm in *Xenopus*. *Development* **126**, 3703–3713 (1999).
10. He, X., Saint-Jeannet, J. P., Woodgett, J. R., Varmus, H. E. & Dawid, I. B. Glycogen synthase kinase-3 and dorsoventral patterning in *Xenopus* embryos. *Nature* **374**, 617–622 (1995).
11. Hsu, S. C., Galceran, J. & Grosschedl, R. Modulation of transcriptional regulation by LEF-1 in response to Wnt-1 signaling and association with β -catenin. *Mol. Cell. Biol.* **18**, 4807–4818 (1998).
12. Yost, C. *et al.* The axis-inducing activity, stability, and subcellular distribution of β -catenin is regulated in *Xenopus* embryos by glycogen synthase kinase 3. *Genes Dev.* **10**, 1443–1454 (1996).
13. Sokol, S. Y. Analysis of Dishevelled signalling pathways during *Xenopus* development. *Curr. Biol.* **6**, 1456–1467 (1996).
14. Slusarski, D. C., Corcos, V. G. & Moon, R. T. Interaction of Wnt and a Frizzled homologue triggers G-protein-linked phosphatidylinositol signalling. *Nature* **390**, 410–413 (1997).
15. Sheldahl, L. C., Park, M., Malbon, C. C. & Moon, R. T. Protein kinase C is differentially stimulated by Wnt and Frizzled homologs in a G-protein-dependent manner. *Curr. Biol.* **9**, 695–698 (1999).
16. Otte, A. P. & Moon, R. T. Protein Kinase C isozymes have distinct roles in neural induction and competence in *Xenopus*. *Cell* **68**, 1021–1029 (1992).
17. Münchberg, S. R. & Steinbeisser, H. The *Xenopus* Ets transcription factor XER81 is a target of the FGF signaling pathway. *Mech. Dev.* **80**, 53–65 (1999).
18. Medina, A. & Steinbeisser, H. Interaction of Frizzled 7 and Dishevelled in *Xenopus*. *Dev. Dyn.* **218**, 671–680 (2000).
19. Deardorff, M. A., Tan, C., Conrad, L. J. & Klein, P. S. Frizzled-8 is expressed in the Spemann organizer and plays a role in early morphogenesis. *Development* **125**, 2687–2700 (1998).
20. Rothbächer, U. *et al.* Dishevelled phosphorylation, subcellular localization and multimerization regulate its role in early embryogenesis. *EMBO J.* **19**, 1010–1022 (2000).
21. Shi, D. L., Goisette, C. & Boucaut, J. C. Expression of Xfz3, a *Xenopus* frizzled family member, is restricted to the early nervous system. *Mech. Dev.* **70**, 35–47 (1998).

Acknowledgements

We thank A. Brändli, I. Dawid, M. Hammerschmidt, P. Klein, M. Kühl, R. Moon, W. Reintsch, U. Rothbächer, R. Rupp, D. L. Shi and S. Sokol for providing plasmids; U. Müller for technical assistance; and S. Cramton for critically reading the manuscript. This work was supported by a pre-doctoral fellowship from the Deutscher Akademischer Austauschdienst (DAAD) to A.M. and by a Deutsche Forschungsgemeinschaft (DFG) grant to R.W.

Correspondence and requests for materials should be addressed to H.S. (e-mail: herbert.steinbeisser@tuebingen.mpg.de).

Gene expression in *Pseudomonas aeruginosa* biofilms

Marvin Whiteley*, M. Gita Banger†, Roger E. Bumgarner‡, Matthew R. Parsek§, Gail M. Teitzel§, Stephen Lory† & E. P. Greenberg*

* Department of Microbiology, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA

† Department of Microbiology and Molecular Genetics, Harvard University, Boston, Massachusetts 02115, USA

‡ Department of Microbiology, University of Washington, Seattle, Washington 98195, USA

§ Department of Civil Engineering, Northwestern University, Chicago, Illinois 60208, USA

Bacteria often adopt a sessile biofilm lifestyle that is resistant to antimicrobial treatment^{1–5}. Opportunistic pathogenic bacteria like *Pseudomonas aeruginosa* can develop persistent infections^{1–3}. To gain insights into the differences between free-living *P. aeruginosa* cells and those in biofilms, and into the mechanisms underlying the resistance of biofilms to antibiotics, we used DNA microarrays. Here we show that, despite the striking differences in lifestyles, only about 1% of genes showed differential

expression in the two growth modes; about 0.5% of genes were activated and about 0.5% were repressed in biofilms. Some of the regulated genes are known to affect antibiotic sensitivity of free-living *P. aeruginosa*. Exposure of biofilms to high levels of the antibiotic tobramycin caused differential expression of 20 genes. We propose that this response is critical for the development of biofilm resistance to tobramycin. Our results show that gene expression in biofilm cells is similar to that in free-living cells but there are a small number of significant differences. Our identification of biofilm-regulated genes points to mechanisms of biofilm resistance to antibiotics.

Many persistent infections are caused by biofilms². One particularly devastating example is the lung infections caused by *Pseudomonas aeruginosa* biofilms in people with the genetic disease cystic fibrosis. Once *P. aeruginosa* colonizes the cystic fibrosis lung it cannot be eradicated by even the most aggressive antibiotic therapy^{2,3,6,7}. Bacteria growing in biofilms possess characteristics distinct from their free-floating or swimming (planktonic) counterparts. Biofilm bacteria are resistant to antimicrobial treatments, and to the immune defences of hosts^{1,2}. Biofilm bacteria form structured communities of cells embedded in an extracellular polymeric (EPS) matrix^{1,4,5}. Biofilm formation by *P. aeruginosa* occurs in discrete steps: surface attachment and multiplication; microcolony formation; and differentiation into mature, structured, antibiotic-resistant communities^{2,8}. The differences in populations of biofilms and planktonic bacteria have led to the hypothesis that there are major differences in gene expression between the two growth modes¹. We used DNA microarray technology to test this hypothesis with *P. aeruginosa* as a model, and to gain insights into possible mechanisms of the ability of biofilm bacteria to withstand exposure to antibiotics.

Gene expression may vary during development, but the first question is how gene expression differs in a mature biofilm and planktonic cells. This is particularly relevant because of the resistance of mature biofilms to antimicrobial treatment. We grew *P. aeruginosa* using continuous-culture techniques with an effort to minimize differences in the conditions to which the bacteria were exposed. Planktonic bacteria were grown in a chemostat near the maximum growth rate for *P. aeruginosa*. Biofilms were also grown in chemostat vessels; however, they were attached to granite pebbles (Fig. 1). The growth medium and dilution rate for the biofilm were the same as for the chemostat culture of planktonic bacteria.

We used a *P. aeruginosa* microarray to compare gene expression in planktonic and biofilm cells⁹. The array contained 5,500 of the predicted 5,570 *P. aeruginosa* genes. Most genes in *P. aeruginosa* strain PAO1 were expressed at levels that allowed an analysis with reasonably small statistical variation (74% of the 5,500 genes arrayed). Few of these genes showed differential expression in

biofilms compared with planktonic cultures (Fig. 2). This result argues strongly against the proposal that existence in a biofilm results in dramatic differences in the overall make-up of bacterial cells¹. However, our analysis averages gene expression in biofilms, which are heterogeneous groups of cells exhibiting different activities^{1,10}. It may be that certain subpopulations in the biofilm had substantially different patterns of gene expression than did the homogeneous planktonic population or most of the metabolically active cells in the biofilm. This could be true for planktonic cultures too.

More interesting than the genes that did not show biofilm-specific gene regulation is the small number of genes (73) that did show differential expression (at least a twofold difference; Table 1). In all, 34 were activated and 39 were repressed in biofilm populations. We validated the array data by analysing expression of several genes by northern blotting and ribonuclease protection assays (see Methods). About 34% of the 73 biofilm-regulated genes code for hypothetical proteins of unknown function. This is slightly lower than the overall percentage of such genes (44%) derived from the genome-sequencing project of *P. aeruginosa*¹¹.

To show that the microarray results applied to biofilms and planktonic cells grown under different conditions, we used ribonuclease protection assays to compare the expression of three genes in a once-flow-through biofilm system⁷ with the expression in a batch culture of planktonic *P. aeruginosa*. The genes were PA2128, gene for probable fimbrial protein; PA1080, gene for flagellar hook protein; and PA0971, *tolA*. We also tested expression of one gene (PA3622, *rpoS*) by using a *lacZ* reporter. Gene expression patterns appeared similar to those obtained in our microarray analysis of *P. aeruginosa* grown in chemostat vessels. For PA2128, we observed 7-fold biofilm repression in tube biofilms versus 16-fold repression in the microarray experiment; for PA1080, 2.5-fold repression versus 2-fold repression; for PA0971, 3-fold induction versus 4-fold induction; and for PA3622, 2-fold repression versus 2.3-fold repression. We do not predict that all of the genes identified in the microarray analysis will show biofilm regulation under all experimental conditions, but these four genes show similar patterns of regulation in at least two different experimental systems.

The most highly activated genes in *P. aeruginosa* biofilms were those of a temperate bacteriophage that is closely related to the filamentous bacteriophage Pf1 (ref. 12). The genome of *P. aeruginosa* PAO1 contains 11 out of 14 Pf1 genes¹¹. We assessed the abundance of Pf1-like phage in the fluid over the biofilms and in the planktonic chemostat culture fluid by using a plaque-formation assay on a Pf1-sensitive strain of *P. aeruginosa*. The induced levels of

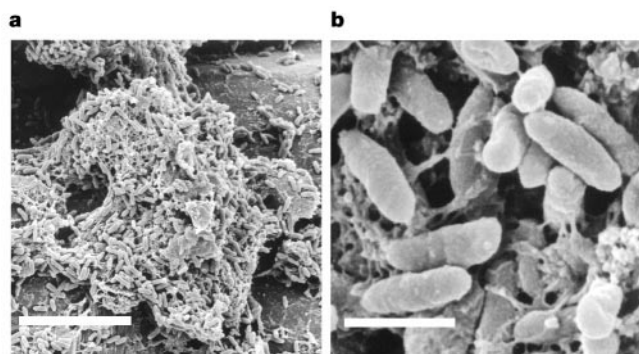


Figure 1 Scanning electron micrographs of a *P. aeruginosa* biofilm on the surface of a granite pebble. **a**, A biofilm on the surface of a pebble. Scale bar, 10 μ m. **b**, High magnification of a biofilm showing rod-shaped *P. aeruginosa* and strings of dehydrated EPS connecting bacterial cells. Scale bar, 1 μ m.

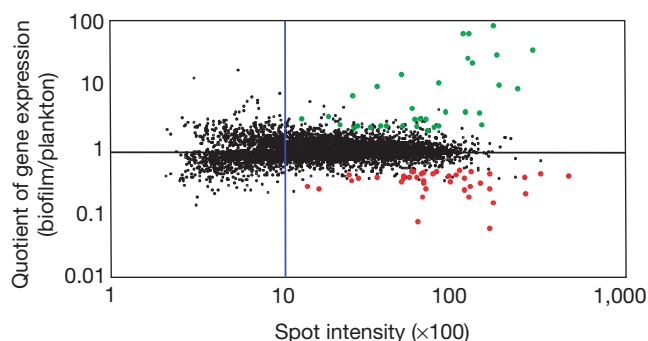


Figure 2 Gene expression and spot intensities of the 5,500 *P. aeruginosa* genes analysed with the microarray. Biofilm-activated genes are in green and biofilm-repressed genes are in red. Black dots represent genes that showed less than twofold regulation, or genes for which the data were not statistically different. Spot intensity refers to average total spot fluorescence (mean of eight independent spots). Spot intensities below 1,000 are shown but were not included in subsequent analyses because of statistical variability³⁰.

Pf1 transcripts in biofilm cells were reflected by a 100–1,000-fold greater abundance of Pf1 in the biofilm reactor than in the plankton reactor. Phage induction might be important for gene transfer within biofilms, it could function in exclusion of other strains of *P. aeruginosa* from a biofilm, or perhaps, as is true for certain temperate bacteriophage in other bacteria, the phage genome contains a gene specifying a toxin^{13,14}. Many phage gene remnants and

transposon-related sequences seem to be induced in *Escherichia coli* biofilms (A. Hay, personal communication).

Genes for synthesis of pili and flagella are repressed in biofilms (Table 1). Pili and flagella are reported to be involved in the initial steps (attachment and microcolony formation) of development of *P. aeruginosa* biofilms⁸. Our results suggest that these appendages may not be required for maintenance of a mature biofilm, but that they are involved in committed steps in biofilm development. Once development has proceeded through these steps, pili and flagella are no longer required.

We asked whether the array analysis provides insights into antibiotic resistance in *P. aeruginosa* biofilms. An analysis of the *P. aeruginosa* genomic sequence revealed more than ten gene clusters seeming to encode efflux pumps of the resistance nodulation division (RND) type¹¹. Activation of some or all of these gene clusters are suggested to be involved in biofilm resistance to antibiotics¹⁵. Our data show that none were induced in the biofilm. However, some of the genes that are activated or repressed in biofilms are known to affect antibiotic sensitivity in *P. aeruginosa*. Cationic antibiotics such as tobramycin and gentamicin bind to the negatively charged lipopolysaccharide (LPS) of the outer membrane^{16,17}, and subsequent transport into *P. aeruginosa* correlates with the level of the transmembrane electrical potential^{18–20}.

The major aminoglycoside-resistance mechanism of *P. aeruginosa* is impermeability of the bacteria to antibiotic entrance^{21,22}. This impermeability involves several factors, including the *tolA* gene product²³ and terminal electron transport proteins^{18,19}. The *tolA* gene product affects LPS structure, resulting in decreased aminoglycoside affinity for the outer membrane. Evidence indicates that *tolA* is an essential *P. aeruginosa* gene, but mutants that under-produce *tolA* are hypersensitive to aminoglycoside antibiotics²³. The

Table 1 Genes differentially expressed in *P. aeruginosa* biofilms

<i>P. aeruginosa</i> ORF	Number	Fold activation (mean $\pm$ s.e.m.)
Bacteriophage genes		
Coat protein B of bacteriophage Pf1	PA0723	83.5 $\pm$ 10.3
Hypothetical protein of bacteriophage Pf1	PA0722	64.2 $\pm$ 5.6
Helix-destabilizing protein of bacteriophage Pf1	PA0720	35.2 $\pm$ 2.7
Hypothetical protein of bacteriophage Pf1	PA0721	26.6 $\pm$ 4.1
Protein of bacteriophage Pf1	PA0718	22.6 $\pm$ 2.9
Hypothetical protein from bacteriophage Pf1	PA0727	14.6 $\pm$ 2.4
Probable coat protein A of bacteriophage Pf1	PA0724	10.1 $\pm$ 0.6
Hypothetical protein of bacteriophage Pf1	PA0725	9.9 $\pm$ 1.1
Hypothetical protein of bacteriophage Pf1	PA0726	8.9 $\pm$ 0.5
Motility and attachment		
Probable fimbrial protein	PA2128	–16.5 $\pm$ 1.5
Pilin protein PilA	PA4525	–6.6 $\pm$ 0.8
Flagellar basal-body rod modification protein FlgD	PA1079	–2.7 $\pm$ 0.3
Probable pili assembly chaperone	PA2129	–2.4 $\pm$ 0.2
Flagellin type B	PA1092	–2.3 $\pm$ 0.3
Flagellar capping protein FlgD	PA1094	–2.1 $\pm$ 0.3
Flagellar hook protein FlgE	PA1080	–2.0 $\pm$ 0.1
Translation		
50S ribosomal protein L28	PA5316	4.4 $\pm$ 0.5
50S ribosomal protein L19	PA3742	2.7 $\pm$ 0.1
50S ribosomal protein L4	PA4262	2.4 $\pm$ 0.2
50S ribosomal protein L18	PA4247	2.3 $\pm$ 0.3
50S ribosomal protein L23	PA4261	2.3 $\pm$ 0.2
30S ribosomal protein S7	PA4267	2.2 $\pm$ 0.3
Translation initiation factor IF-2	PA4744	2.1 $\pm$ 0.1
Ribosome modulation factor	PA3049	–5.3 $\pm$ 0.7
ATP-binding protease component ClpA	PA2620	–2.1 $\pm$ 0.1
Metabolism		
Urease β subunit	PA4867	63.1 $\pm$ 8.1
Ferredoxin (4Fe-4S)	PA0362	2.9 $\pm$ 0.3
Lipoate-protein ligase B	PA3997	2.8 $\pm$ 0.4
Glycerol-3-phosphate dehydrogenase	PA3584	–4.1 $\pm$ 0.3
Cytochrome c oxidase, subunit III	PA0108	–2.9 $\pm$ 0.3
Cytochrome c oxidase, subunit II	PA0105	–2.9 $\pm$ 0.2
Cytochrome c oxidase, subunit I	PA0106	–2.7 $\pm$ 0.2
Leucine dehydrogenase	PA3418	–2.5 $\pm$ 0.2
Membrane proteins or secretion		
Translocation protein TatB	PA5069	6.9 $\pm$ 1.4
TolA protein	PA0971	3.9 $\pm$ 0.4
Translocation protein TatA	PA5068	2.4 $\pm$ 0.2
Outer membrane lipoprotein OmlA	PA4765	2.4 $\pm$ 0.7
Probable porin	PA3038	–3.5 $\pm$ 0.5
Type III secretion central regulator	PA1710	–2.5 $\pm$ 0.3
Probable sodium:solute symporter	PA3234	–2.3 $\pm$ 0.1
Regulation		
Probable transcriptional regulator	PA2547	3.1 $\pm$ 0.1
σ -factor RpoH	PA0376	2.3 $\pm$ 0.3
σ -factor RpoS	PA3622	–2.3 $\pm$ 0.3
Probable two-component response regulator	PA4296	–2.2 $\pm$ 0.2
Other		
Rod shape-determining protein MreC	PA4480	3.1 $\pm$ 0.5
Probable DNA-binding protein	PA5348	–4.6 $\pm$ 0.4
Probable glycosyl hydrolase	PA2160	–2.3 $\pm$ 0.2
Methylated-DNA-protein-cysteine methyltransferase	PA0995	–2.1 $\pm$ 0.2

Conserved hypothetical

PA0990, 4.0 $\pm$ 0.4; PA0579, 3.3 $\pm$ 0.7; PA2971, 2.4 $\pm$ 0.2
PA3785, –3.9 $\pm$ 0.5; PA3235, –3.2 $\pm$ 0.2; PA4738, –3.1 $\pm$ 0.2; PA2621, –3.0 $\pm$ 0.3;
PA1017, –2.7 $\pm$ 0.2; PA0588, 2.6 $\pm$ 0.3; PA1533, –2.6 $\pm$ 0.3; PA0587, –2.3 $\pm$ 0.2

Hypothetical

PA1870, 29.7 $\pm$ 1.2; PA3884, 11.1 $\pm$ 1.6; PA3231, 3.8 $\pm$ 0.4; PA0714, 2.5 $\pm$ 0.3;
PA1372, 2.5 $\pm$ 0.3
PA3411, –12.8 $\pm$ 1.7; PA1676, –5.2 $\pm$ 0.6; PA1830, –3.9 $\pm$ 0.9; PA4638, –3.7 $\pm$ 0.8;
PA1244, –3.1 $\pm$ 0.4; PA1855, –2.9 $\pm$ 0.2; PA4607, –2.6 $\pm$ 0.3; PA4661, –2.4 $\pm$ 0.2;
PA3922, –2.1 $\pm$ 0.2

Positive values represent activation in biofilms. Negative values represent repression.

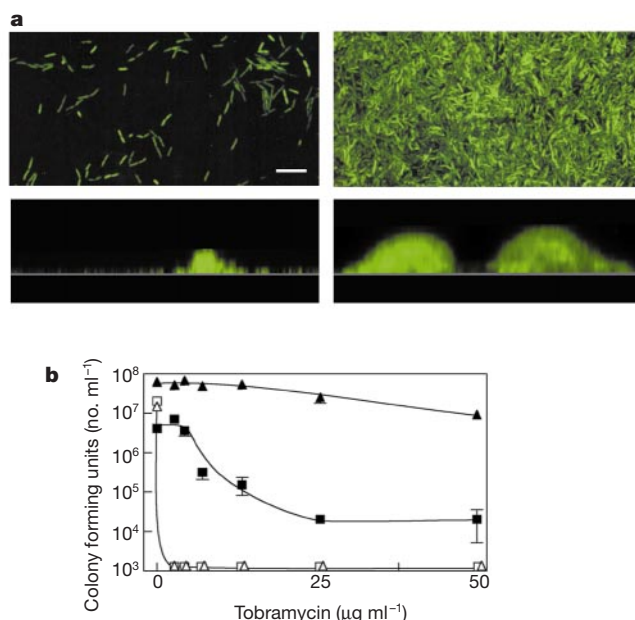


Figure 3 Comparison of wild-type and *rpoS*-mutant *P. aeruginosa* biofilms.

a, Epifluorescent photomicrographs (top) and scanning confocal sagittal reconstructions (bottom) of wild-type (left) and *rpoS*-mutant (right) *P. aeruginosa* containing a GFP-expression plasmid. Scale bar, 5 μ m. **b**, Survival of cells in wild-type and *rpoS*-mutant biofilms after treatment with the antibiotic tobramycin. Open symbols, planktonic cultures; filled symbols, biofilms; squares, parent strain PA01; triangles, *rpoS* mutant. Error bars show s.e.m. and are too small to be seen in most cases. At tobramycin concentrations above 5 μ g ml^{–1} there were <2 viable cells ml^{–1} in planktonic cultures. Complementation of the *rpoS* mutant with the plasmid pMW105 containing a functional *rpoS* gene²⁶ restored biofilm formation rates and tobramycin survival to levels similar to wild-type levels (data not shown).

tolA gene was activated in *P. aeruginosa* biofilms (Table 1). This clearly could contribute to the resistance of the biofilms to aminoglycosides. The cytochrome *c* oxidase genes were repressed. Cytochrome *c* oxidase is the terminal electron acceptor during aerobic growth, and repression of cytochrome *c* oxidase should decrease sensitivity of *P. aeruginosa* to aminoglycoside antibiotics^{18,19}. These are just two examples of genes that may be involved in biofilm resistance to one class of antibiotics. Many other genes in Table 1 might be involved as well (for example, the porin genes). The genes coding for proteins of unknown function are particularly interesting candidates as antibiotic-resistance factors. If such a gene is involved in antibiotic resistance, functional studies may reveal previously unknown biofilm resistance mechanisms.

If the microarray experiment (Table 1) has identified genes important in biofilm development or antibiotic resistance, then mutants defective in some of these genes should show aberrant biofilm structure and antibiotic sensitivity. We studied a mutant defective in one of the biofilm-regulated genes, *rpoS*. The *rpoS* gene codes for an RNA polymerase σ subunit and it influences transcription of other *P. aeruginosa* genes¹⁵. Previous studies indicated that an *rpoS*-deletion mutant of *P. aeruginosa* was hypervirulent in a mouse model²⁴, and that *rpoS* may serve some role in biofilm development²⁵. The *rpoS* gene was repressed in our array experiment (Table 1) and in the once-flow-through tube biofilm reactor system.

To examine the involvement of *rpoS* in biofilm development, we grew the wild-type strain and an isogenic *rpoS* mutant²⁶, both tagged with a *gfp* plasmid (which expresses green fluorescent protein (GFP)), in flow-cell reactors. The *P. aeruginosa* biofilms that developed in the flow cells were examined by scanning confocal laser microscopy (Fig. 3a). Within 4 h, differences in the *rpoS* mutant and wild-type biofilms were evident. The mutant had attached to and covered much more of the glass surface. In a quantitative microtitre dish assay⁸, biofilm formation of the parent strain was 38% of the *rpoS* mutant. After 24 h, the mutant biofilm had matured and large structured groups of bacteria were evident. The wild-type biofilm showed smaller structures (Fig. 3a), and after a further incubation the wild-type biofilm remained thinner than the mutant biofilm. This is consistent with a previous examination of our *rpoS* mutant: a 6-day-old *rpoS* mutant biofilm showed a mean thickness of 17 μm , and the thickness of the parent biofilm was 6 μm (ref. 25).

To assess whether *rpoS* influenced the susceptibility of *P. aeruginosa* biofilms to antibiotics, we treated biofilms of the *P. aeruginosa* wild type and *rpoS* mutant with increasing amounts of the aminoglycoside antibiotic tobramycin. Tobramycin is a front-line antibiotic used in the treatment of *P. aeruginosa* infections²⁷.

Biofilms of the *rpoS* mutant were much more resistant to killing by tobramycin than were wild-type *P. aeruginosa* biofilms (Fig. 3b). These results indicate that the *rpoS* gene is important for biofilm formation and in susceptibility to the antibiotic tobramycin. These studies also indicate that our array analysis identified a gene important for biofilm development and antibiotic sensitivity. It seems likely that other genes identified in the microarray analysis will be important in biofilm development and antibiotic susceptibility.

We hypothesize that existence in a biofilm induces moderate levels of resistance to all antimicrobial treatments. This could afford cells in a biofilm the opportunity to respond to an antibiotic by inducing genes more specific to that antibiotic. We compared biofilms exposed to tobramycin with untreated biofilms. Consistent with our hypothesis, 20 genes were differentially expressed in tobramycin-treated biofilms (Table 2): 14 were activated and 6 were repressed by tobramycin (seven times the minimum inhibitory concentration for planktonic cells). Of these 20 genes, 12 were classified as genes coding for hypothetical proteins of unknown function. As expected, treatment with tobramycin, which causes errors in protein synthesis, seemed to induce a stress response, with activation of *dnaK* and *groES* genes, for example. Tobramycin strongly induced several genes coding for hypothetical proteins. It also induced two probable efflux systems (a probable non-RND drug efflux system and a P-type ATPase). These are candidate tobramycin-resistance loci. Four genes that were expressed at a higher level in biofilms than in plankton were repressed by tobramycin treatment of biofilms, and two that were repressed in biofilms were activated in tobramycin-treated biofilms.

This initial microarray analysis of *P. aeruginosa* biofilms shows that, on average, gene expression in biofilm cells is remarkably similar to gene expression in planktonic cells maintained under similar environmental conditions. We identified only 73 genes that were differentially expressed in biofilms. Our studies of one biofilm-regulated gene, *rpoS*, showed it to be involved in the morphology and antibiotic sensitivity of biofilms. Further studies will reveal which of the biofilm-induced genes are activated under different environmental conditions, and which are important in biofilm biology. Such a subset of genes will be of great use in the development of rapid screens for agents that block biofilm maintenance. □

Methods

Planktonic and biofilm culture conditions

For array experiments, *P. aeruginosa* strain PAO1 was grown at 37 °C with aeration in chemostat vessels (100 ml of medium, dilution rate 0.2 h⁻¹). The growth medium consisted of 0.5% NH₄Cl, 0.25% NaCl, 0.025% Mg SO₄, 7H₂O, 0.015% KH₂PO₄, 1.5% MOPS buffer at pH 7.0, non-chelated trace elements and 0.015% casamino acids. For biofilms, the chemostat vessel contained 100 g of sterilized granite pebbles (~70 pebbles). Vessels were inoculated with 10⁶ *P. aeruginosa* cells. For biofilm cultures, bacteria were allowed to attach to the pebbles for 24 h before the flow of medium was initiated, which was at a high rate (100 ml h⁻¹). After 4 h at the high flow rate, >99% of the unattached bacteria were washed away. The flow was then decreased to a dilution rate identical to that in the planktonic chemostat. After 5 d, cell numbers in the biofilm and planktonic chemostats were similar (10¹⁰ cells) as determined by standard plate-counting techniques. To remove biofilm bacteria for plate counting, pebbles were vortexed for 2 min in 5 ml phosphate-buffered saline. A small planktonic population was present in the biofilm reactors throughout the experiment, but never represented more than 0.1% of the total chemostat population.

To validate the microarray experiments in one way, biofilms were grown in a once-flow-through system in silicone tubing as previously described⁷, except that 20% Luria Bertani (LB) broth was used as the growth medium. For studies of biofilm development, *P. aeruginosa* containing the *gfp* plasmid pMRP-1 were grown in flow-cell reactors and examined by scanning confocal microscopy²⁸. For tobramycin-susceptibility studies, we grew biofilms using a spinning disk reactor and treated them with tobramycin as outlined elsewhere²⁹.

To assess the influence of tobramycin on biofilm gene expression, we added the antibiotic (5 $\mu\text{g ml}^{-1}$) to the influent medium after 4 d of biofilm growth. After 24 h in the presence of tobramycin, we removed biofilms and isolated the RNA. A concentration of 5 $\mu\text{g ml}^{-1}$ of tobramycin is about seven times the minimum inhibitory concentration for planktonic *P. aeruginosa* strain PAO1. Standard plate counting indicated that 5 $\mu\text{g ml}^{-1}$

Table 2 Genes regulated by tobramycin in *P. aeruginosa* biofilms

<i>P. aeruginosa</i> ORF	Number	Fold activation (mean $\pm$ s.e.m.)
Ribosome modulation factor	PA3049	24.7 $\pm$ 11.4
Hypothetical protein	PA4326	23.2 $\pm$ 7.5
Hypothetical protein	PA2703	19.3 $\pm$ 3.7
Probable DNA-binding protein	PA5348	18.1 $\pm$ 4.9
Hypothetical protein	PA1110	8.7 $\pm$ 1.7
Conserved hypothetical protein	PA3463	8.3 $\pm$ 3.2
GroES	PA4386	5.8 $\pm$ 1.1
Conserved hypothetical protein	PA3785	5.0 $\pm$ 1.3
Probable drug-efflux protein	PA1541	4.7 $\pm$ 1.4
Probable metal-transporting P-type ATPase	PA3920	3.6 $\pm$ 1.1
Probable transcriptional regulator	PA3574	3.3 $\pm$ 0.6
Conserved hypothetical protein	PA2498	2.9 $\pm$ 0.6
DnaK	PA4761	2.7 $\pm$ 0.4
Conserved hypothetical protein	PA3819	2.1 $\pm$ 0.3
Hypothetical protein	PA1893	-2.4 $\pm$ 0.3
Hypothetical protein of bacteriophage Pf1	PA0725	-3.1 $\pm$ 0.5
Urease β subunit	PA4867	-3.7 $\pm$ 1.1
Hypothetical protein of bacteriophage Pf1	PA0721	-3.9 $\pm$ 1.0
Hypothetical protein	PA1870	-6.4 $\pm$ 1.5
Hypothetical protein	PA3884	-14.7 $\pm$ 5.7

Positive values represent activation by tobramycin. Negative values represent repression.

tobramycin did not measurably affect biofilm cell numbers. However, this level of tobramycin effectively killed planktonically cultured *P. aeruginosa* (Fig. 3b).

Scanning electron microscopy

Biofilm samples were prepared for scanning electron microscopy by fixation with 2.5% glutaraldehyde and staining with 1% osmium tetroxide. The samples were dehydrated in ethanol and hexamethyldisilazane, air dried, mounted on aluminium stubs, and sputter coated with gold and palladium (60:40). Imaging was conducted with a Hitachi S-4000 scanning electron microscope. Scanning confocal microscopy was performed as previously described²⁸.

Microarray analysis

We isolated RNA from planktonic and biofilm bacteria using the Trizol reagent (Life Technologies). Planktonic cells were formed into pellets by centrifugation (7,000 g for 10 min) at 4 °C. For isolation of biofilm RNA, rocks containing biofilm bacteria were rinsed with sterile medium and suspended in Trizol reagent. After vortexing for 1 min, the rocks were removed and the remaining cell material was sonicated for 10 s. Insoluble material was removed by centrifugation and RNA was isolated as described above. Contaminating DNA was eliminated by DNase treatment, and RNA was isolated by phenol-chloroform extraction and precipitation with LiCl. Complementary DNA probes were produced from RNA with random primers (NSNSNSNSNS) and Cy5-dCTP or Cy3-dCTP (Amersham) according to previously described procedures³⁰. To avoid complications associated with Cy5-dCTP and Cy3-dCTP incorporation rates into resulting cDNA, each RNA comparison was performed with both dye combinations on separate microarrays. The microarrays were glass microscope slides containing representative gene-specific DNA fragments from 5,500 of the estimated 5,570 open reading frames (ORFs) of *P. aeruginosa*⁹. The microarrays were printed with a Generation II Array printer (Molecular Dynamics), and the hybridized microarrays were imaged with a Generation II scanning confocal fluorescent microscope (Molecular Dynamics). A full description of the microarray will be published later. Statistical analysis of microarray data was performed with previously described computer software³⁰.

The data in Tables 1 and 2 represent results of two independent experiments (means of eight individual comparisons). *Pseudomonas aeruginosa* ORF numbers and homologies were obtained from the *Pseudomonas* Genome Project (<http://www.pseudomonas.com>). Classifications are based on those described by Stover *et al.*¹¹. Northern blot analysis and ribonuclease protection assays were performed as outlined by the manufacturer (Ambion) to verify microarray data for four activated and repressed genes. These techniques revealed fold regulations of 75 ± 10, 3.6 ± 0.7, -25 ± 5, and -3.0 ± 0.7 for PA4867, PA0971, PA2128, and PA1080, respectively.

Gene expression in once-flow-through tube biofilms (5 d old) was compared with gene expression in mid-logarithmic-phase planktonic cultures grown in LB broth. Analysis was either by quantitative ribonuclease protection assays or by monitoring β-galactosidase activity in a strain with a *lacZ* reporter, as indicated.

Received 9 April; accepted 18 September 2001.

- Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. & Lappin-Scott, H. M. Microbial biofilms. *Annu. Rev. Microbiol.* **49**, 711–745 (1995).
- Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318–1322 (1999).
- Burns, J. L., Ramsey, B. W. & Smith, A. L. Clinical manifestations and treatment of pulmonary infections in cystic fibrosis. *Adv. Pediatr. Infect. Dis.* **8**, 53–66 (1993).
- DeBeer, D., Stoodley, P. & Lewandowski, Z. Liquid flow in heterogenous biofilms. *Biotech. Bioeng.* **44**, 636–641 (1994).
- Lawrence, J. R., Korber, D. R., Hoyle, B. D., Costerton, J. W. & Caldwell, D. E. Optical sectioning of microbial biofilms. *J. Bacteriol.* **173**, 6558–6567 (1991).
- Hoiby, N. Antibiotic therapy for chronic infection of pseudomonas in the lung. *Annu. Rev. Med.* **44**, 1–10 (1993).
- Singh, P. K. *et al.* Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **407**, 762–764 (2000).

- O'Toole, G. A. & Kolter, R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* **30**, 295–304 (1998).
- Bangera, M. G., Ichikawa, J. K., Marx, C. & Lory, S. in *Am. Soc. Microbiol. 100th Gen. Meeting* (ASM Press, Los Angeles, 2000).
- Costerton, J. W. *et al.* Biofilms, the customized microniche. *J. Bacteriol.* **176**, 2137–2142 (1994).
- Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
- Hill, D. F., Short, N. J., Perham, R. N. & Petersen, G. B. DNA sequence of the filamentous bacteriophage Pf1. *J. Mol. Biol.* **218**, 349–364 (1991).
- Zabriskie, J. B. Viral-induced bacterial toxins. *Annu. Rev. Med.* **17**, 337–350 (1966).
- Johnson, J. A., Morris, J. G. & Kaper, J. B. Gene encoding zonula occludens toxin (*zot*) does not occur independently from cholera enterotoxin genes (*ctx*) in *Vibrio cholerae*. *J. Clin. Microbiol.* **31**, 732–733 (1993).
- Greenberg, E. P. Bacterial genomics: Pump up the versatility. *Nature* **406**, 947–948 (2000).
- Hancock, R. E. W. Alterations in outer membrane permeability. *Annu. Rev. Microbiol.* **38**, 237–264 (1984).
- Nikaido, H. & Vaara, M. Molecular basis of bacterial outer membrane permeability. *Microbiol. Rev.* **49**, 872–878 (1985).
- Bryan, L. E. & Kwan, S. Roles of ribosomal binding, membrane potential, and electron transport in bacterial uptake of streptomycin and gentamicin. *Antimicrob. Agents Chemother.* **23**, 835–845 (1983).
- Bryan, L. E., Nicas, T., Holloway, B. W. & Crowther, C. Aminoglycoside-resistant mutation of *Pseudomonas aeruginosa* defective in cytochrome c552 and nitrate reductase. *Antimicrob. Agents Chemother.* **17**, 71–79 (1980).
- Damper, P. D. & Epstein, W. Role of membrane potential in bacterial resistance to aminoglycoside antibiotics. *Antimicrob. Agents Chemother.* **20**, 803–808 (1981).
- Bryan, L. E., Haraphongse, R. & Elzen, H. M. V. D. Gentamicin resistance in clinical-isolates of *Pseudomonas aeruginosa* associated with diminished gentamicin accumulation and no detectable enzymatic modification. *J. Antibiot. (Tokyo)* **29**, 743–753 (1976).
- MacLeod, D. L. *et al.* Aminoglycoside-resistance mechanisms for cystic fibrosis *Pseudomonas aeruginosa* isolates are unchanged by long-term, intermittent, inhaled tobramycin treatment. *J. Infect. Dis.* **181**, 1180–1184 (2000).
- Rivera, M., Hancock, R. E., Sawyer, J. G., Haug, A. & McGroarty, E. J. Enhanced binding of polycationic antibiotics to lipopolysaccharide from an aminoglycoside-supersusceptible, *tolA* mutant strain of *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **32**, 649–655 (1988).
- Suh, S. J. *et al.* Effect of *rpoS* mutation on the stress response and expression of virulence factors in *Pseudomonas aeruginosa*. *J. Bacteriol.* **181**, 3890–3897 (1999).
- Heydorn, A. *et al.* Experimental reproducibility in flow-chamber biofilms. *Microbiology* **146**, 2409–2415 (2000).
- Whiteley, M., Parsek, M. R. & Greenberg, E. P. Regulation of quorum sensing by RpoS in *Pseudomonas aeruginosa*. *J. Bacteriol.* **182**, 4356–4360 (2000).
- Hoiby, N. *Pseudomonas in Cystic Fibrosis: Past, Present and Future* 1–25 (Cystic Fibrosis Trust, Berlin, 1998).
- Davies, D. G. *et al.* The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* **280**, 295–298 (1998).
- Hentzer, M. *et al.* Alginate overproduction affects *Pseudomonas aeruginosa* biofilm structure and function. *J. Bacteriol.* **183**, 5395–5401 (2001).
- Ichikawa, J. K. *et al.* Interaction of *Pseudomonas aeruginosa* with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs. *Proc. Natl Acad. Sci. USA* **97**, 9659–9664 (2000).

Acknowledgements

We thank T. Moninger for assistance with electron microscopy and A. Hay for providing information on *E. coli* biofilm gene expression in advance of publication. Supported by a grant from the National Institutes of Health (E.P.G.), a New Technologies Grant from the Cystic Fibrosis Foundation (S.L.), and a grant from the Procter & Gamble Company (E.P.G.). M.W. was supported by a National Science Foundation Research Training Grant and a US Public Health Service Training Grant. G.M.T. is supported by a National Science Foundation Center Grant.

Correspondence and requests for materials should be addressed to E.P.G. (e-mail: everett-greenberg@uiowa.edu).

Oligonucleotides on view

Buying, hybridizing, and measuring oligonucleotides and DNA.

OligoBuyer

Tangerine Technologies

www.tangerinetechnology.com

Online sourcing for oligonucleotides

OligoBuyer software makes it possible to buy complex strands of DNA online from a range of suppliers. Part of the Tangerine EpiCenter suite of software, OligoBuyer was developed to provide biopharmaceutical research organizations with a web-based system for buying complex biomolecules from multiple suppliers in a single interface. EpiCenter applications also allow 'e-procurement' of oligos, peptides, antibodies and microarray chips.

Pan Oligo Sets

MWG-Biotech

www.mwg-biotech.com

Kit for producing DNA chips

MWG-Biotech's new oligonucleotide sets enable customers to produce their own oligonucleotide microarrays. The sets are based on the Pan Array series of catalogue DNA chips and on the company's proprietary HPSF oligonucleotide technology. The sets contain all nucleotides needed for spotting on a glass slide, together with buffer solutions and protocol. If required, MWG-Biotech will also aliquot to customer specification. The service is complemented by optional process optimized glass slides and MWG-Biotech's arrayer and scanner series.

HM-4000 Multidizer

UVP, Inc.

www.uvp.com

Ready to rock and roll

There are two chambers in the HM-4000 Multidizer hybridization oven, for hybridization and blotting procedures requiring different settings. Each compartment operates independently to create the conditions needed for cDNA library screenings, primer synthesis



Chamber made: the two-in-one HM-4000.

and all nucleic acid hybridizations. The lower chamber features temperature controls, variable rotation speed and multiple bottle size options, with rock-and-roll action available with an optional rocker plate. The upper chamber has a roller action to operate shaker tray or optional orbital tray.

GeneLinkerGold version 1.1

Molecular Mining Corporation

www.molecularmining.com

An upgrade for expression analysis software

GeneLinker Gold version 1.1 gene expression analysis software is designed to be easy to use, comprehensive, and affordable. Functions introduced in Version 1.1 include Principal Components Analysis (PCA), Self Organizing Maps (SOMs) and the Jarvis-Patrick clustering feature. Data importation, on-line help, and visualization tools are as introduced in

version 1.0. GeneLinker Gold is licensed in perpetuity to the end user, rather than on an annual basis. Software upgrades and tech support are free of charge for one year, after which the user can choose to stay with this version or to purchase upgrades and support as required.

GeneAmp PCR

Applied Biosystems

appliedbiosystems.com

Amplify up to 10 kb of human genomic DNA

The GeneAmp High Fidelity PCR system is formulated with a blend of AmpliTaq DNA Polymerase and a new thermostable proof-reading enzyme. This combination is claimed to increase fidelity threefold and double the yield of Taq DNA Polymerase. GeneAmp is recommended for cloning, site-directed mutagenesis, and other applications requiring accurate nucleotide incorporation. The High Fidelity system amplifies up to 10 kb of human genomic DNA and will operate with a wide range of template samples, including mammalian genomic DNA, artificial chromosomes, plasmids, and post-reverse transcription.

BioMini spectrophotometer

Shimadzu

www.shimadzu.com

Spectrophotometry for nucleic acids

Shimadzu's BioMini UV-Vis spectrophotometer is aimed at the life sciences market, chiefly for nucleic acid analysis, protein quantitation and spectrum measurement. The nucleic acid module provides quantitation methods for double and single stranded DNA, RNA and standard Warburg-Christian. DNA melting calculations provide information for optimization of PCK and hybridization procedures.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

**DIVERSE
SMALL
MOLECULES
READY FOR
SCREENING**

**10K-50K
Compounds
in Plates**

- Next Day Delivery*
- Competitively Priced

*(Some Restrictions Apply)

Chembridge Corporation • 800.964.6143
858.451.7400 • 858.451.7401 fax • E-mail: chem@chembridge.com

www.chembridge.com

**Polyclonal Antibodies
& Custom Peptides**

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

ResGen™
Invitrogen Corporation

U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our web site at
www.resgen.com

CUSTOMISED
immunoassays for
research and production

BioGenes
Gesellschaft für Biopolymere mbH

Internet www.biogenes.de
Berlin • Germany
Phone +49 (0)30-65 76 23 96

Proteomics technology: Character references

technology feature

In the *Odyssey*, Homer's hero has his hands full when he faces Proteus. The demigod challenges Odysseus by transforming himself into a lion, a boar, a serpent, a wave and finally a tree. In proteomics, scientists trying to discern the nature of proteins face an equally formidable challenge, because protein data are as mutable as Proteus. Protein levels in different cell types change constantly as they are upregulated, downregulated, cleaved and phosphorylated.

Because protein information, unlike DNA, is not static in the cell, scientists must follow Odysseus' lead. They will have to be resourceful, especially as the tools used in today's high-throughput environment still bear the stamp of an earlier era when one protein at a time was the standard.

The 2D gel used to separate individual proteins from complex mixtures dates back to the mid-1970s. Mass spectrometry, which identifies proteins by weight once they are isolated, has been around since the First World War. And industrial robots,

used to usher the proteins through the intermediate steps that separate these two techniques, date back to the 1960s. Most venerable of all is the century-old separations technique of chromatography.

Fortunately for scientists aiming for widespread protein characterization in the wake of the triumph of genome sequencing, a series of improvements in mass spectrometry and 2D-gel technology is readying these tools for the task that lies ahead. Chromatography was modernized in the 1970s with the invention of high-pressure pumps, the addition of multiple columns and improved packing materials for columns, leading to its modern incarnation as high-performance liquid chromatography, or HPLC, a workhorse in many life-sciences labs.

Established scientific-equipment companies are also working to integrate more steps of the overall proteomics workflow into fewer pieces of equipment. And many start-up companies are looking for ways to enhance or supplant parts of the established proteomics process.



The mass spectrometer is key to proteomics.

JPS

Although there are many different methods emerging — from mapping all the proteins in a single organism to describing the multitude of interactions experienced by proteins during their lifespan — the general technique of isolating and identifying the many proteins in different cell types remains central.

MASS SPECTROMETRY

Mix and match

Mass spectrometry represents a worldwide market worth US\$1 billion a year, with about a third of that dedicated to machines especially suited for proteomics. The system uses three components — an ionization source, an analyser and a detector. Users have at least two choices for each component. Assorted pairings offer different advantages — some combinations are more suited

to proteomics, whereas others lend themselves more to small-molecule analysis. And some combinations will integrate with other proteomics equipment such as liquid chromatography.

Companies are tending to make their new machines more versatile, more automated and more compatible with other proteomics equipment — but, in general, the more choices offered by one

machine, the higher the price tag.

The choices begin where the process starts — ionization sources. Ionization gives the sample an electric charge. The widely used MALDI (matrix-assisted laser-desorption/ionization) uses solid samples, and produces ions of large and small molecules. Electrospray ionization (ESI) is used less often in proteomics. It ionizes liquid samples and is most often used for peptides and small molecules. It can be directly coupled to liquid chromatography systems.

For analysis, time-of-flight (TOF) is most frequently used with MALDI, whereas ESI is usually coupled to quadrupole or ion-trap analysers. Quadrupole machines are considered low-performance instruments compared with MALDI-TOF, but they only cost about a third as much. Ion-trap analysers are also modest performers, but they are robust and easier to look after than the other types, and are even more modestly priced.

Finally, there are two kinds of mass spectrometer — MS and MS/MS. MS is the faster, easier-to-operate option. But, in addition to generating a spectrum of the sample, MS/MS can take some of the ions that have been separated and measured, fragment them further, and then generate spectra of those parts. This allows users to discern which amino acids the peptides contain, and, in some cases, can identify the sequence of these amino acids within the peptide.



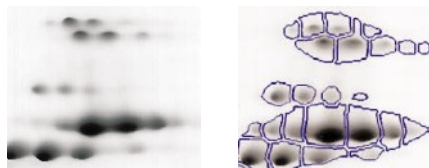
Integration: speeding analysis.

There are several possible starting points for protein identification. But the most well-travelled route into proteomics starts with a sample in a 2D gel being fed into an electrophoresis machine. This is followed by either automatic or manual picking and excision of the protein spots of interest, which are then fed into a mass spectrometer (see 'Mix and match', previous page).

Celia Caulcott, who heads an effort by the UK's Biotechnology and Biological Sciences Research Council to develop new proteomics technologies, says that, despite a lot of R&D, traditional techniques for protein identification still stand. "The gels still seem to be the pre-eminent way people want to do things," she says. Beguiling techniques such as protein arrays, which could supplant gels if successful, have yet to prove they can be viable both scientifically and commercially, she says.

Joakim Rodin, director for proteomics R&D at Amersham Biosciences, a biotech-equipment company based in Uppsala, Sweden, agrees that the gel system, although not the easiest thing to work with, has yet to be supplanted. "It's still a lot of work running the gels," he says. But improvements in capacity, such as the company's Ettan Dalt II system, allows researchers to run up to 12 gels in parallel with more reproducibility and sensitivity.

And the gels themselves have improved,



Identifying spots on gels can be time consuming.

he says. They are getting bigger, so more sample can be loaded, which improves the detection of low-abundance proteins. 'Zoom' gels have also been developed with ever-narrowing pH ranges, which give better resolution as well as higher sensitivity.

Fluorescent labelling is also getting better, he says. Differential-expression analysis using difference gel electrophoresis, developed at Carnegie Mellon University, allows up to three samples to be run simultaneously on a single gel using cyanine-dye chemistry. This should let researchers detect protein differences between normal and cancerous tissues on the same gel. The method also allows multiplexing of gels, which significantly increases throughput, reproducibility and accuracy. Multiple gels provide comparative analysis and accurate measurement of differential protein expression. Although the handling and analysis of 2D gels have improved dramatically, Rodin notes that complementary techniques, such as X-ray

crystallography, are needed to resolve the whole proteome.

Fortunately, the next stage of the proteomics pipeline, handling the intermediate steps between electrophoresis and mass spectrometry, is becoming easier. Picking the protein spots off the gels, then digesting them into peptide fragments used to be two separate, manual tasks. Now they are becoming automated and are being integrated into the workflow (see 'Multiple choice', below). But improving and combining individual components can be challenging, says Steve Martin, director of Applied Biosystems' Proteomics Research Center in Framingham, Massachusetts. For example, increasing the capacity of one instrument without accounting for the additional need for throughput in others can actually result in bottlenecks, he says.

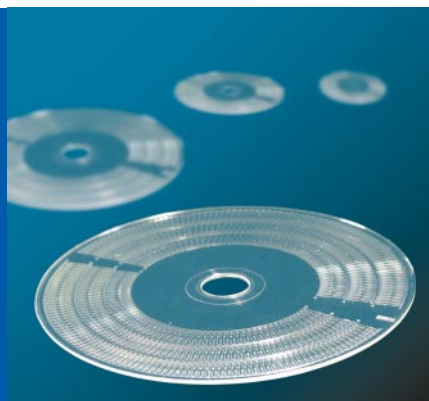
Three commercial — and by today's standards, integrated — systems are made by Amersham Biosciences, Genomic Solutions in Ann Arbor, Michigan, and Bio-Rad in Hercules, California. Their basic components are similar — they all use robotic sample-preparation, 2D-gel electrophoresis, excision of spots, labelling, and ionization and analysis of the peptide fragments by mass spectrometry. In these systems, data generated from all the instruments are presented in a user-friendly graphical interface.

AUTOMATION

Until recently, characterizing proteins was done slowly. But with the many candidates in the newly sequenced genomes crying out for attention, and the lure of complex protein assemblies beckoning, labs are gearing up to look at many proteins simultaneously.

The key to making such a system work lies in replacing error-prone humans with spot-picking robots, guided by cameras and sophisticated image-analysis software. The Australian company ARRM has a system that excises spots from gels or polyvinylidene fluoride membranes and places them in a 96-well plate for subsequent proteolysis.

Genetix, of New Milton, UK, uses a line of sample-preparation, gel-spotting and spot-excision units. Soon these will be joined by a machine to prepare MALDI samples automatically, thus helping to integrate raw samples and mass fingerprints. Genetix is also getting into chip arrays and yeast two-hybrid systems, two automated ways of looking at protein interactions. Other major players in lab automation are Amersham Biosciences in Uppsala,



Lab-on-a-CD systems from Gyros.

in Uppsala. Gyros has updated an idea from the 1970s by engraving microscale channels and mixing chambers on a compact disc. Centrifugal force and controlled surface chemistry are used to regulate the flow of liquid through the CD. Despite the small size of the system, the price tag will probably ensure that it will mainly be used by big pharmaceutical companies or 'protein factories' rather than small independent labs.

Multiple choice

Sweden, Bio-Rad in Hercules, California, and Genomic Solutions in Ann Arbor, Michigan.

Large Scale Biology in Germantown, Maryland, and Oxford Glycoscience in Cambridge, UK, aim to automate the entire protein-discovery process in humidity-controlled, robot-populated buildings. Here massive amounts of samples would travel through the pipeline from gel to mass spectrometer and data.

But harking back to the idea that small is beautiful, another school of thought sees a nano future for the science in 'lab-on-a-chip' technologies such as those of Caliper, of Fremont, California, and Gyros

These stations are quite expensive but, just as core facilities for genome sequencing sprang up once the equipment came of age, the same is likely to happen with protein characterization. This should ensure that smaller academic and commercial labs will share in the advance of knowledge. And smaller labs might still be able to automate individual steps, such as spot picking or digestion, finding new ways to integrate steps that might be overlooked in larger, more streamlined organizations.

Alternatives for eliminating, rather than integrating, such steps are also emerging. One fairly new strategy involves transferring the gel to a membrane made of polyvinylidene difluoride (PVDF), then probing the membrane directly with mass spectrometry. This bypasses the spot-cutting step between electrophoresis and mass spectrometry.

Improvements also extend to mundane but essential items such as stains. Coomassie blue, a staple in most labs, can interfere with the digestion of gel spots by trypsin, so new stains such as zinc imidazole and noncovalent fluorescent SYPRO dyes, which do not have this limitation, are being introduced.

Mass-spectrometry output

It was not until the early 1990s that the mass spectrometers, now virtually essential components in the proteomics pipeline,

could be used to analyse proteins.

Mass spectrometry relies on the fact that a substance carrying a net electric charge — an ion — can be made to move in a predictable way in an electromagnetic field. Ions are sorted by their charge-to-mass ratio, and from these a 'mass fingerprint' of the sample can be derived. Software, such as the University of California's Prospector package, can then be used to match the fingerprint to a protein database such as Amos Bairoch's Swiss-Prot (see 'Setting standards', overleaf).

In earlier models, excessive ionization energies would blast delicate molecules such as DNA and proteins into indecipherable particles. But innovations using a matrix such as MALDI, which protects the sample by modulating the ionizing laser beam, have helped to overcome this limitation.

Nevertheless, the technique still has its limits. A mass fingerprint will not be enough for identification if the protein is not registered in a database, or if post-translational modifications have changed its observed mass from the predicted value. In these instances, more information can be obtained from secondary protein fragments by re-routing the ions from the first analysis down a second channel and then analysing these fragments with the spectrometer. Of course, more complete databases will also help. And pairing mass

spectrometry with other techniques, such as some kinds of protein-detector chip (see 'Alternative approaches', below) may make the method even more useful.

Future challenges

Automating and integrating the protein-characterization process is a good start, but there is no simple way forward. Although effective with adequate sample sizes, automated processes in general are not effective with very small amounts (less than 10 femtomoles of material).

It is hard enough to describe a single protein in a particular state. But things get even more difficult when trying to



Leroy Hood (right) and Ruedi Aebersold.

CHIPS

Proteins lack DNA's copying ability and do not readily undergo amplification, making separation and fractionation more important — especially for small amounts of proteins. And the inherent complexity and diversity of proteins makes a viable protein array an even more difficult goal. But the need to process proteins en masse is so urgent that heroic efforts are under way to develop a workable protein chip.

Leading the field at present are designs based on antibodies tethered to a solid surface. Large Scale Biology in Germantown, Maryland, and Biosite Diagnostics in San Diego, California, are developing an array of antibodies against 2,000–5,000 protein targets from the former's human protein index database. Biosite will use its omniconal phage display technology to generate high-affinity antibodies against the targets. The companies hope the system will be available in the second half of 2002.

But an inherent drawback of antibody chips — or any protein chip, for that matter — is the destructive effect of proteases that may be lurking in the analyte mixture. "You have to use protease inhibitors if you're sampling microdissected tissue," says Lance Liotta of the US National Cancer Institute's Center for Cancer Research, who invents tools for proteomics and has surveyed the existing technology. "Process the tissue, lyse it, stain it and pray

Alternative approaches

that these manipulations don't affect the 3D state of the protein."

Perhaps the biggest challenge is the accurate quantification of low-abundance protein. The faint signal of a protein of interest may easily be swamped by the much higher concentrations of other surrounding proteins.

Ciphergen in Fremont, California, is selling a device that helps scientists to detect low-abundance proteins. The company's chip uses specific surface chemistries to affinity-capture minute quantities of proteins. "A peak in one sample but not the other says a variation exists, but you still have to figure out what it is," says Mike Baldwin, a chemist at the University of California, San Francisco. "It's an interesting approach, but not mainstream proteomics — at least, not yet."

Another recent quantitative protein-expression and -identification technique using mass spectrometry is isotope-coded affinity tagging (ICAT), a kind of labelling invented by Ruedi Aebersold at the Institute for Systems Biology in Seattle. The start-up company Sense Proteomic, based in Cambridge, UK, is trying to use smaller numbers of mounted proteins to assay for suspected protein-protein interactions such as those known to play a role in toxicity.

Other chip approaches towards proteomics include atomic-force microscopy, aptamer libraries and biosensors.

characterize thousands of proteins active at any time in various parts of the cell. Michael Washburn and Dirk Wolters at Syngenta Agricultural Discovery Institute in San Diego and John Yates at the Scripps Research Institute in La Jolla, California, have devised a system to separate and identify 1,484 proteins from the proteome of the yeast *Saccharomyces cerevisiae* (see *Nature Biotechnol.* **19**, 242–247; 2001). But that relatively low number in the humble yeast doesn't begin to reveal the complexity in humans. For example, there are a thousand or more proteins involved in the G-signalling pathway, which regulates everything from the most basic activities of the cell (division, motility) to the most specialized

ones (secretion, electrical excitability).

Perhaps the biggest hurdle is not in designing the equipment but in the conceptual realm. Researchers might know individual elements in a signal cascade, understand something about their function, and perhaps even have obtained their structure. But, explains Ehud Isacoff, a biophysicist at the University of California, Berkeley, scientists are still encumbered by a bias to view the overall picture as if it were made up of discrete events, with one protein handing a signal to another sequentially, in a series of 'stills'.

What is really happening in the cell, Isacoff continues, "is that proteins are very localized, and dock against one another very precisely in assemblies, and signalling happens by molecular motions that propagate from one subunit to another". New methodologies and systems of notation must be devised to describe these things, and a new breed of student has to be recruited who can think about them as concrete objects with specific structures and interactions.

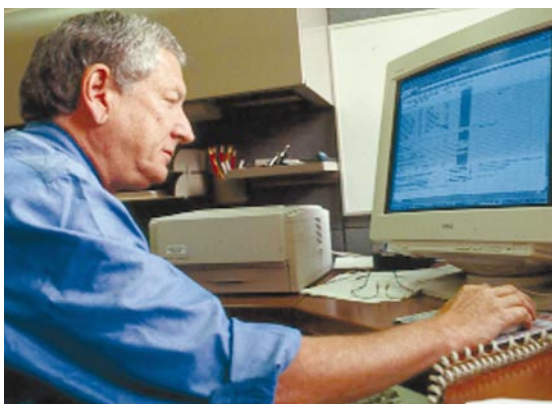
In fact, these needs are being recognized and the integrative effort is under way on several fronts. Leroy Hood's Institute for Systems Biology in Seattle has been in existence since early last year (see *Nature* **407**, 828–829; 2000), and Al Gilman's Alliance for Cell Signalling at Dallas set up shop a year ago (see *Nature* **407**, 7; 2000).

They aim at a holistic understanding of the cell in all of its pathways and interactions. New methodology — and, perhaps, improved equipment — may emerge from such efforts.

And a Clinical Proteomics Initiative, under the aegis of the US National Institutes of Health, started seeking grant applications last month. One of its key elements will be the antibody consortium, says Lance Liotta of the National Cancer Institute and one of those engineering the enterprise. This will be modelled on the open-access but industry-supported SNP consortium that is mapping simple genetic variations. Support — both in terms of finance and willingness to donate antibodies — from industrial and academic groups is very enthusiastic, says Liotta. The consortium's ultimate goal is to develop and make available arrays of every antibody and every ligand in existence.

Other aspects of the NIH initiative are looking for new approaches to existing techniques. However, it's unlikely that any new technology will completely replace an old one. Instead, innovations arising from the initiative will probably occur alongside the stalwarts of electrophoresis, mass spectrometry and chromatography — further complicating the ever-changing face of proteomics.

Potter Wickware is a science writer based in San Francisco; Paul Smaglik is *Naturejobs* editor.



Al Gilman: seeking the cell's secrets.

SOFTWARE

In the realm of software and databases, there is a real opportunity for integration, but instead developers have tended to go off in their own directions. Great strides have been made in areas such as image analysis and peak-picking tools for mass spectrometry with software packages including Tycho, Melanie and Quest. Software developed by Nonlinear Dynamics of Newcastle upon Tyne, UK, aids in spot detection on gels and also helps in quantitative analysis of those spots once they are picked. Major

equipment manufacturers Amersham Biosciences and PerkinElmer have already signed on to bundle this program, called Progenesis, with some of their instruments. But, according to Patsy Babbitt, a protein informaticist at the University of California, San Francisco, the software side is fragmented. "It's a big problem," Babbitt says.

Organizations such as the Bio-Ontologies Consortium aim to clarify the picture with

standards and nomenclature, but perhaps what is lacking are new ways of thinking about the information generated in proteomics — classical bioinformatics is built around pattern-matching algorithms.

Tony Pawson and Chris Hogue at the University of Toronto have been thinking about the informatics side of proteomics. They have developed the Biomolecular Interaction Network Database (BIND), which indexes interactions between DNA, RNA, proteins and small molecules, as well as temporal and compartmental information. As BIND's content grows, "we'll be the GenBank of interactions", predicts Francis Ouellette, of the University of British Columbia in Vancouver, one of the resource's developers. Other databases for proteomics include the Database of Interacting Proteins at the University of California, Los Angeles, Large Scale Biology's Human Protein Index and Atlas Base, by the San Diego company Accelrys, which contains protein structures.



Tony Pawson.



Chris Hogue.

GENERAL TECHNOLOGY

Company	Product/activity	Location	URL
Advion BioSciences	Has an ESI chip for MS manufactured by Intellisense. Contract LC company.	Ithaca, New York	www.advion.com
Affibody	Uses proprietary protein to study protein–protein interactions. Associated with Gyros.	Stockholm, Sweden	www.affibody.com
ARRM	Robotics, associated with Hochstrasser and Genetix.	Adelaide, Australia	www.rrm.com
Biacore	Ligand fishing; surface plasmon resonance for real-time detection and monitoring of biomolecular binding events.	Uppsala, Sweden	www.biacore.se/proteomics
BioRobotics	Robotics for arrays and colony picking.	Cambridge, UK	www.biorobotics.com
Borealis Biosciences	Protein identification, affinity chromatography.	Toronto, Canada	www.protein-affinity.com
Caliper	Microfluidics-based 'lab-on-a-chip'. Capillary-based 'sipper chip' for continuous screening assays. Fluorogenic, electrophoretic mobility, including kinases and phosphatases. To come are calcium flux and membrane potential. Collaborates with Structural GenomiX and Agilent.	Mountain View, California	www.calipertech.com
Caprion Pharmaceuticals	CellCarta organelle purification-based identification of rare proteins.	Montreal, Canada	www.caprion.com
Ciphergen	Protein-chip reader, chip software, chip arrays (hydrophilic/hydrophobic, anion/cation exchange, immobilized metal), SELDI chip, protein-chip arrays for affinity capture.	Fremont, California	www.ciphergen.com
Cytomyx	Provides proteomics services from sample preparation through to MS.	Cambridge, UK	www.cytomyx.com
Europroteome	Works on epithelial cancers.	Hennigsdorf, Germany	www.europroteome.com
Genetix	GelPix spot excision, array products, robotics.	New Milton, UK	www.genetix.com
Genomic Solutions	Investigator proteomic system, fully integrated solution for identifying and characterizing proteins. ProGest sample preparation station.	Ann Arbor, Michigan	www.genomicsolutions.com
Gyros	Nanoscale 'lab-on-a-CD' for MS.	Uppsala, Sweden	www.gyrosmicro.com
Ingeny	On the verge of introducing an automated 2D-gel system.	Goes, Netherlands	www.ingeny.com
Integrative Proteomics	Uses Bruker XR, NMR, MS equipment for functional proteomics.	Toronto, Canada	www.integrativeproteomics.com
Large Scale Biology	Produces therapeutic proteins in plants. Is developing an antibody chip with Biosite 2D-gel factory.	Vacaville, California	www.lsb.com
LumiCyte	Protein biomarker profiles for drug development.	Fremont, California	www.lumicyte.com
MediChem	Protein expression, crystallization, structure, through subsidiary companies.	Woodridge, Illinois	www.medicchem.com/services/proteomics.htm
Myriad Genetics	ProNet yeast two-hybrid protein interaction system, ProSpec MS-based drug discovery.	Salt Lake City, Utah	www.myriad.com
NextGen Sciences	Platform technologies in proteomics, transcriptomics and genomics.	Huntingdon, UK	www.nextgensciences.com
Oxford GlycoSciences	2D gels, developing ICAT/MALDI-TOF system.	Oxford, UK	www.ogs.com
Pepscan Systems	Protein–protein interactions based on epitopes. epitope mapping, combinatorial peptide arrays and peptidomimetics. Lead development service.	Lelystad, Netherlands	www.pepscan.nl
Phylos	Antibody-based high-throughput 'HIP' chip; ProFusion selection technology based on protein fused to its own mRNA.	Lexington, Massachusetts	www.phylos.com
Protagen	Phosphorylation analysis.	Bochum, Germany	www.protagen.de
Proteomic Solutions	Proteomics service company.	St Marcel, France	www.proteomicsolutions.fr
Protein Pathways	Pathway analysis in mammals and microbes.	Los Angeles	www.proteinpathways.com
Proteome Sciences	Identifies protein markers in human disease targets. US subsidiary is Intronn LLC.	Cobham, Surrey, UK	www.proteome.co.uk
Proteome Systems	Glycosylation analysis, 2D-gel excision equipment, chemical printer for electroblotting gels to membranes, Axima CFR MS machine.	Sydney, Australia	www.proteomesystems.com
ProteoSys	Analyses modifications, quantifies down to sub-attomolar range.	Mainz, Germany	www.proteosys.com
Rigel	Oncology proteins, ubiquitin ligase.	South San Francisco, CA	www.rigel.com
Roche	Proteomics research centre.	Basel, Switzerland	www.roche.com
Sense Proteomic	COVET functional protein array.	Cambridge, UK	www.sensetherapeutic.com
Sensium Technologies	Uses biosensors to detect low abundance proteins, proteomics subsidiary of Aurora/Helix BioPharma.	Edmonton, Canada	www.helixbiopharma.com
SomaLogic	Aptamer arrays, partner with Celera.	Boulder, Colorado	www.somallogic.com
Syrrx	Structural proteomics, proteomics discovery platform with Eli Lilly.	San Diego, California	www.syrrx.com
Virtek	Desktop array reader can be used for 2D gels.	Ontario, Canada	www.virtek.ca
WITA Proteomics	High-resolution 2D-gel platform, protein identification. Alliance with Pharmagene.	Tetlow, Germany	www.wita-proteomics.com
Zymark	Robots for screening, liquid handling, plate management.	Hopkinton, Massachusetts	www.zymark.com
Zyomyx	Protein-chip development.	Hayward, California	www.zyomyx.com

MASS SPECTROMETRY

Company	Products	Location	URL
Agilent Technologies	Ion-trap MS machines.	Palo Alto, California	www.agilent.com/chem
Amersham Biosciences	Ettan DALT II MALDI-TOF MS.	Uppsala, Sweden	www.amershambiosciences.com
Analytica of Branford	Prototype MS developer.	Branford, Connecticut	www.aob.com
Applied Biosystems	API series (with MDS Sciex) and Voyager MS machines.	Foster City, California	www.appliedbiosystems.com
Bio-Rad	ProteomeWorks System, with MicroMass.	Hercules, California	www.proteomeworkssystem.com
Bruker Daltonics	Ultraflex 3000 TOF/TOF tandem MS, SNAP, biotools, ProteinScope software.	Bremen, Germany	www.daltonics.bruker.com
Hitachi Instruments	Many kinds of MS, LC & analytical equipment.	Yokohama, Japan	www.hii-hitachi.com
MDS Sciex	API series LC/MS systems.	Toronto, Canada	www.sciex.com
Micromass	ProteomeWorks System marketed by Bio-Rad.	Manchester, UK	www.micromass.co.uk
ThermoFinnigan	Quadrupole, TOF and ion-trap MS machines.	San Jose, California	www.thermofinnigan.com

REAGENT SUPPLIERS

Company	Products	Location	URL
Amicon	Immobilon transfer membranes division of Millipore.	Bedford, Massachusetts	www.millipore.com/lifescience/products.nsf/docs/amicon
BioWhittaker	Reagents, part of Cambrex Life Sciences. Has lab service centre in North Brunswick, New Jersey.	East Rutherford, New Jersey	www.cambrex.com
CBS Scientific	Reagents for gels.	Del Mar, California	www.cbssci.com
Crescent Chemical	Electrophoresis and LC chemicals.	New York	www.creschem.com
Fermentas	Reagents, protein standards.	Vilnius, Lithuania	www.fermentas.com
IBA	Protein expression and purification with STREP-tag technology.	Göttingen, Germany	www.iba-go.de
ICN Biomedicals	Reagents, radioisotopes.	Costa Mesa, California	www.icnbiomed.com
Invitrogen	Reagents, membranes, filters.	Paisley, UK	www.invitrogen.com
Jule Biotechnologies	Precast gels.	New Haven, Connecticut	hometown.aol.com/precastgel/index.htm
Molecular Probes	Stains.	Leiden, Netherlands	www.probes.com
Novagen	Protein purification and analysis products.	Madison, Wisconsin	www.novagen.com
Pierce	Chemicals, reagents, trypsin, antibodies, cytokines.	Rockford, Illinois	www.piercenet.com
Promega	Expression cloning system.	Madison, Wisconsin	www.promega.com
Qiagen	Protein expression and purification, also has Biorobot 8000 for purifying proteins.	Hilden, Germany	www.qiagen.com
R. Shadel	Gel boxes, combs, gaskets, spacers.	San Francisco	www.shadel.com
Schleicher & Schuell	Filters, membranes.	Dassel, Germany	www.s-und-s.de
Sigma-Aldrich	Reagents, purified trypsin.	St Louis, Missouri	www.sigma-aldrich.com
Stratagene	<i>In vitro</i> mutagenesis kit.	La Jolla, California	www.stratagene.com
Zaxis	Precast gels.	Hudson, Ohio	www.zaxis-inc.com

SOFTWARE

Company	Products	Location	URL
Accelrys	Gene Atlas, Atlas Base.	San Diego, California	www.accelrys.com
AxCell Biosciences	GLD ligand identification, ProChart pathway identification.	Newtown, Pennsylvania	www.axcellbio.com
BIND	Database of pathways and interactions.	Toronto, Canada	www.bind.ca
BioBridge Computing	PIUMS (Protein Identification Using Mass Spec), Pepex peak-picking software, launch date November 2001.	Lund, Sweden	www.biobridge.se
Cognia	Protein catabolism database (ubiquitin-dependent turnover for regulation of protein levels, structural homology elements for substrate recognition by E3 enzymes). Also markets Biobase, from University of Aarhus, Denmark.	New York	www.cognia.com
Compugen	Z3 @d-gel image analysis system, Protocall MS analysis tools.	Tel Aviv, Israel	www.cgen.com
Deltagen	DeltaBase, functional annotation DB, phenotype and knockout information.	Redwood City, California	www.deltagen.com
Hybrigenics	PIMRider, protein pathways and interactions software.	Paris, France	pim.hybrigenics.com
Incyte Genomics	Life Express protein database.	Palo Alto, California	www.incyte.com
Large Scale Biology	Human protein index, molecular anatomy pathways, molecular effects of disease databases.	Germantown, Maryland	www.lsb.com
Matrix Science	Mascot software searches sequence databases with MS signatures.	London, UK	www.matrixscience.com
Media Cybernetics	ArrayPro image-analysis software.	Silver Spring, Maryland	www.mediacy.com
Micromass	ProteomeWorks, MassLynx, MetaboLynx MS software.	Manchester, UK	www.micromass.co.uk
Nonlinear Dynamics	Phoretix, Progenesis gel image-analysis software.	Newcastle upon Tyne, UK	www.nonlinear.com
Protein Pathways/UCLA	DIP, database of interacting proteins.	Los Angeles, California	dip.doe-mbi.ucla.edu
Proteometrics	Profound search engine, Enterprise M/Z MS analysis software, Knexus, Radars informatics platforms, BIOML mark-up language.	New York	www.proteometrics.com
Scimagix	Image-analysis software.	Redwood Shores, California	www.scimagix.com
Tripos	Data mining/management tools.	St Louis, Missouri	www.tripos.com

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Wells (4954)

European Manager:

Nevin Bayouni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayouni (4978)

Scandinavia: Silje Opstrup (4994)

Spain/ Portugal: Leonie Wells (4954)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

MG Ichigaya Building (6F),
19-1 Haraikatanmachi,
Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejp.com

Japan Manager:

Kate Cowan

naturejobs

Getting a handle on data

Bioinformatics arose as a field where people wrote their own software or taught themselves how to use the tools that were publicly available. As genomics data piled up, so, too, did programs to mine them. Scientists wanting to learn how to use the latest tool to explore the most recent data have been playing catch-up ever since.

But the educational obstacles appear to be getting even more daunting. More kinds of data — such as protein expression and structure — are accumulating, and programmers are now writing new tools to characterize them and integrate them with the old data. As a result, experienced bioinformaticians must constantly upgrade their skills, and neophytes must work harder to enter the field. More and better training for both groups seems necessary.

Several approaches show promise. Workshops are a start, but their quality tends to be inconsistent (see page 6). There is also evidence that the quantity of such workshops is lacking. One recent workshop at a bioinformatics institute in Europe that had 25 slots received 170 applicants (see page 4).

One alternative? Create a better ongoing dialogue between the people who write the software and those who use it. The US National Institutes of Health (NIH) has launched just such a scheme — bioinformatics experts plan to train small scientific groups from each of the NIH's 22 institutes (see page 7). But this approach has its limitations — for now, the programme applies only to scientists who conduct intramural research, not those who are funded to perform research outside NIH facilities. But if it is adopted by a wider community, such ongoing dialogues between developers and users could create better software, and result in more people who can exploit all the available features.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Europe scrambles for
bioinformatics skills p4

Quality-control problems
afflict American courses p6

Training gets priority
treatment in Japan p8

WWW.NATUREJOBS.COM

Career centre
Information on the scientific
job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

A permanent education

As data analysis shifts towards protein expression and structure determination, demands for standardization and ease of access are leaving bioinformaticians scrambling to keep their skills up to date, says Helen Gavaghan.

GSK



David Searls: ambitious scientists will find exciting work in bioinformatics.

Andy Brass (right) co-founded the popular bioinformatics course at Manchester.

Anyone who has taken apart a comparatively simple machine, such as a vacuum cleaner, in an effort to figure out why it won't work — or how it is supposed to work — knows just how difficult it is to fit all the bits back together. There is usually some disobliging piece left over that turns out to be crucial.

So pity the biologists. Having energetically deconstructed organisms ranging from fruitflies to humans into strings of bases, they are now reassembling them in a way meant to elucidate life, death, disease and evolution at every scale from organelle, through to cells, organs and organisms.

Enter bioinformatics, the merger of computing and biology, and a subject that David Searls, head of bioinformatics training at pharmaceuticals company GlaxoSmithKline and one of the founders of the field, likes to call the midwife of the post-genomic era. Ambitious scientists in this area who can cross disciplines will have exciting work for a long time to come, Searls says.

But getting and keeping jobs in bioinformatics means gaining an ever-shifting skill set. Practitioners

must, of course, learn how to analyse genomes comparatively as more organisms are sequenced. And now bioinformaticians must also tackle microarray data. In the future, as even more data from ever more kinds of sources enter repositories, bioinformaticians will need to work within standards for everything from terminology to software.

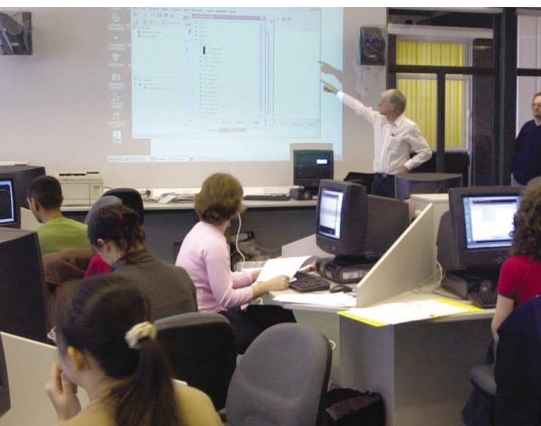
Although a clutch of scientists such as Searls first saw the promise of combining biology with computing in the late 1980s, biologists, even five or six years ago, did not appreciate just how central to their discipline computers would become.

Now there is no doubt, and there is a clamouring at the gates of universities offering bioinformatics courses. In Britain, the University of Manchester's masters course in bioinformatics is routinely oversubscribed. And this year, when it added a distance-learning masters in the subject, it did not even need to advertise for students.

In just a decade, bioinformatics has evolved from a curiosity into a fully fledged academic discipline and industrial activity. It now has all the usual accoutrements, including dedicated journals, societies, international meetings, textbooks, divisions in industry, dedicated chairs and academic courses (at post- not undergraduate level).

FRIENDLY EXPRESSIONS

Bioinformatics is also a discipline on the move, which means there is a constant need to update skills. "Until a year or so ago, we were sequence obsessed," says Alan Robinson, team leader of the industrial group at the European Bioinformatics Institute (EBI) in Hinxton, near Cambridge, UK. Acutely conscious of the need to broaden his knowledge, Robinson is taking a sabbatical in Michael Ashburner's laboratory at the University of Cambridge (until this year Ashburner also headed bioinformatics research at the EBI). Robinson plans to use the sabbatical to brush up on his 'omics' (such as



UNIV. MANCHESTER

proteomics and functional genomics) and to study the potential of computer models of biological processes.

The technology driving many of the developments in the 'omics' world is microarrays. Andy Brass, who helped to establish bioinformatics as a discipline at the University of Manchester, says that in the future it will be as routine to put expression data into public databases as it currently is to deposit sequences.

Already, the EBI has made gene expression central to its mission. At the beginning of October, Janet Thornton, a professor of protein-structure analysis at University College London, took up a five-year post as head of the EBI with responsibility for research. Not surprisingly, gene expression and training in the latest bioinformatics techniques are high on her agenda.

This month, too, the EBI began accepting expression data and ran a workshop on the analysis and informatics of DNA-array gene-expression data. "When the institute offered this workshop, 170 people applied for 25 places," Robinson says.

SHIFT TO STANDARDS

In the long term, the dominant issue will be standards. This could encompass tasks as basic as developing or consulting tables of synonyms to ensure everyone is calling genes and proteins by the same name, or developing open-source, standard chunks of mark-up code, or establishing standard interfaces so that scientists can write their own programs confident that they will work on different databases.

One even more ambitious international standardization programme is the genome ontology

collaboration. This seeks to assign common codes to genes and proteins of a particular function, cellular location or biological role, thus enabling searches across different databases. As a result, a particular apoptosis regulator, say, in the database of one model organism, could be located in another.

Although the thought of developing standards for search tools, database access or terminology, or of learning what these standards are, may not set the pulse racing, this is the way of the future. Without standards, there is the risk that crucial biological data will be left languishing in some inaccessible database — and a biologist who dismantled a piece of biological machinery may never reassemble it into working order.

Helen Gavaghan is a freelance writer based in West Yorkshire.



Janet Thornton: bioinformatics training is top of her priority list.

Web links

European Bioinformatics Institute ♦ www.ebi.ac.uk
 Gene Ontology Consortium ♦ www.geneontology.org
 Bioinformatics resource ♦ www.hgmp.mrc.ac.uk/CCP11/index.jsp
 Bioperl Project ♦ www.bioperl.org
 BioJava Project ♦ www.biojava.org
 Java tutorial ♦ java.sun.com/docs/books/tutorial
 University of Manchester bioinformatics ♦ bioinf.man.ac.uk
 UCL biochemistry group ♦ www.biochem.ucl.ac.uk/bsm
 Nature's genome gateway ♦ www.nature.com/genomics

Training options abound

Oversubscription of bioinformatics courses is not unusual. So a fair degree of self-reliance, including scouring the Internet for tutorials in bioinformatics — or its constituent parts of biology and aspects of computing — is essential for anyone wanting to learn about or keep up to date with the subject.

Fortunately, there are many

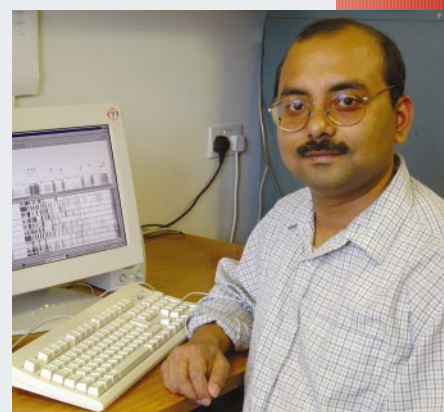
excellent websites for willing autodidacts, says Alan Robinson, team leader of the industrial group at the European Bioinformatics Institute (EBI) in Hinxton, near Cambridge, UK. But non-biologists should consider going along to their friendly local biology laboratory, he says.

David Searls, head of bioinformatics training at GlaxoSmithKline, adds that they should develop their skills using real data, and not work in an 'ivory tower' of idealized problems.

Biologists and non-biologists alike would benefit from participating in ventures such as Bioperl and BioJava, says Robinson, which aim to develop open sources of software for bioinformatics tasks. Apart from enhancing skills, and learning about which solutions work in the real world, participants become

known to one another and the networking helps with job searches.

Nevertheless, some formal training is still helpful. For Debashis Rana, now working as a project scientist on *Brassica* at the John Innes Centre in Norwich, UK, studying bioinformatics using distance-learning modules for the masters course at the University of Manchester is proving invaluable. His biology background is strong — he studied agricultural science as an undergraduate at the College of Agriculture in Vidya-Bharati, India, and then made deep forays into genetics during his PhD and postdoctoral work in the United States. Now he is conscious of the need to figure out how to store, manage and explore efficiently the data that four different institutes in Britain are producing on *Brassica*.



Experience pays: Debashis Rana

One thing is certain, whether people choose a full-time course, distance learning, on-the-job problem solving or web exploration, bioinformatics is moving fast, and those who don't move with the flow and update their skills will miss out on the biological harvest of genomics.

H.G.



Data demystified: Alan Robinson teaching.

Training in a hybrid discipline

Courses to teach bioinformatics are starting to spring up all over North America. But the interdisciplinary nature of the subject means that there is a severe lack of experienced instructors, so the quality may vary between programmes, warns Potter Wickware.

Ararity five years ago, dedicated bioinformatics training programmes are now proliferating in North America. The International Society for Computational Biology lists more than 30 degree programmes in bioinformatics, covering 13 states and provinces in North America, and many more workshops and short courses. A total of 95 offerings worldwide for this marriage between biology and computer science are listed at Rockefeller University's website. Most of today's programmes tend to emphasize computer science but, with the emergence of functional genomics and proteomics, a

shift in emphasis is now needed.

Francis Ouellette, director of the Bioinformatics Core Facility at the Centre for Molecular Medicine and Therapeutics in Vancouver, Canada, sees a need for greater flexibility in university rules if bioinformatics training is to improve and flourish.

"In Canada, and probably in the United States, undergraduate programmes in bioinformatics try to do a full core in biochem and a full core in computer science. But these overloaded programmes don't leave enough flexibility for the other courses that an undergraduate ought to be able to enjoy," Ouellette says. "Even then, hybrid undergraduate bioinformatics degrees have incomplete biology and not enough computer science."

Universities should start bending their rules to make

computer science more available to the biologists, Ouellette explains. "They should make it easy for biologists to get computer science courses, and vice versa, and even bring in other sciences as well," he says.

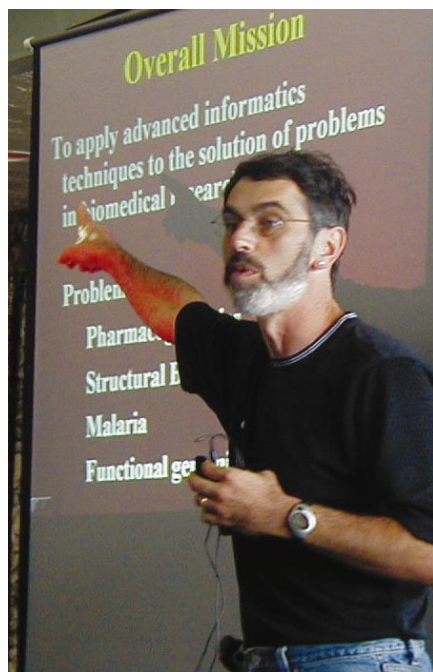
Another problem is the shortage of faculty with full bioinformatics credentials. "Many are biochemists who know a little computer science, or vice versa," Ouellette says. "It's a necessary compromise, and the hope is that there'll be enough cross-pollination for the programmes to work."

Why aren't there more senior bioinformatics investigators and instructors? As with so much in science, funding is key. "Only now are the funding agencies catching on that these projects are fundable and that it's worth attracting people to work in bioinformatics," Ouellette says.

GOING PAST HALFWAY

He believes that graduate degree programmes represent the only sustainable model for bioinformatics training. "I've been to biotech companies in Canada where their directors of bioinformatics turn out to be people who have taken my workshops, and I know how little they know," he says. "The workshops are getting people halfway there — but you're not going to get the innovators, the people who are thinking of new tools and new ways of thinking about the data, from a short course or workshop."

Russ Altman, director of the biomedical informatics training programme at Stanford University and president of the International Society for Computational Biology, thinks there's a critical distinction between programmes that train users and those that train tool-builders. "Tool users should take a short course, then use the tools on the job and continue getting ongoing training," he says. Tool builders, by contrast, need more training because they



Russ Altman: computer science must recognize the challenges being issued by biology.



Francis Ouellette sees graduate degree programmes as key to successful bioinformatics training.

must anticipate the problems that users will report when the tools begin to fall short.

Looking ahead, Altman sees deepening connections between bioinformatics, clinical medicine and databases, which will require an awareness of clinical issues by those who are being trained today.

"In basic science there will be more focus on complex systems, mathematical simulation and integration of data and knowledge," says Altman. "Structure, especially as a milestone on the way to functional prediction, will continue to grow in importance, as will interest in simulation."

As for the scope and quality of today's training programmes, Altman faults them on the same grounds as Ouellette. "Computer science, statistics and related disciplines are slow to recognize the challenges that biology is tossing out," he says. "Databases need to be integrated, new problems in statistics are piling up, and the computer engineers need to start looking at distributed systems in biology. The computer science and statistics departments need to start collaborating more with the biologists."

Tomorrow's practitioners will also need to be sophisticated about informed consent and patient privacy issues, given the increasing interconnectedness of databases and the need for large clinical trials.

The data arising from protein analysis will need skilled interpretation.

VARIAN



NIH starts training

Although the US National Institutes of Health (NIH) has spawned many key bioinformatics tools, it has only recently launched a concerted effort to teach people how to use them.

This initiative has been made difficult because the types of data — and the tools to mine them — keep changing. In addition, top bioinformaticians are often tempted to leave academia for lucrative posts in industry.

Outside the NIH, foundations are being laid through a body of the institute that is only a few months old. The NIH's Center for Bioinformatics and Computational Biology (CBCB), established in May, has so far awarded grants to three US universities to set up pre-doctoral training programmes — the University of California, San Diego; Washington University, St Louis, Missouri; and Stanford University. Each is getting funds to establish up to three teaching posts, with

the understanding that they try to bolster them with support staff.

Jim Cassatt, the CBCB's acting director, notes that recruiting for such positions is difficult because of competition with industry.

Inside the NIH, the National Center for Biotechnology Information (NCBI) plans to train groups at each of the NIH's 22 institutes in the finer points of the NCBI's bioinformatics software. Those people would, in turn, educate others working at the NIH.

The NCBI's David Wheeler, who is heading the training, says that most people who use bioinformatics software only scratch the surface of what it can do.

So far, five people have been trained, with four more in progress. If the programme is successful on the NIH campus, the NCBI may consider expanding it to help extramural researchers funded by the agency, but no such plans have yet been approved. **Paul Smaglik**

"The social-ethical side is lagging," Altman says. He is pressing for progress on this point as part of the Pharmacogenetics Knowledge Base project, a database project supported by the National Institutes of Health, and in the Stanford Informatics programme, where PhD candidates are focusing on ethical, legal and social issues in bioinformatics. But he also notes that passing new laws and regulations must proceed at a deliberate pace, while the technological advances that make them necessary are rushing forward unchecked. ■

Potter Wickware is a science writer in San Francisco.

Web links

Canadian Bioinformatics Workshops ♦ www.bioinformatics.ca

Centre for Molecular Medicine and Therapeutics

♦ www.cmmt.ubc.ca/ouellette

Pathogenomics Project ♦ www.pathogenomics.bc.ca

Biomolecular Interaction Network Database ♦ www.bind.ca

Course list at Rockefeller University

♦ linkage.rockefeller.edu/wli/bioinfocourse

International Society for Computational Biology ♦ www.iscib.org



Playing catch-up



Japan's government is belatedly realizing that it needs to increase funding for training in bioinformatics, says Robert Triendl. But lack of specialists in the field could hinder the country's efforts.

Researchers at the Computational Biology Research Center (top) are joining those at Kyoto University's bioinformatics centre in Japan's drive towards computational biology.

Well-trained bioinformatics specialists in Japan are not just rare — they are virtually non-existent. This is partly because of a lack of formal education in the subject, and the problem is systemic.

With little formal recognition of bioinformatics as a field, graduate departments have until recently allocated only a limited number of students to existing bioinformatics teachers. The government recognized the need for more bioinformaticians as it scaled up the country's genomics efforts.

This year, Japan's Ministry of Education, Science, Sports and Culture started to upgrade bioinformatics education at national universities by creating additional staff positions and funding both undergraduate courses and graduate-level informatics training.

Kyoto University's new bioinformatics centre is a product of this new policy. The university is Japan's leading academic centre for bioinformatics research, but until a few months ago all its bioinformatics activities were concentrated in just one laboratory: Minoru Kanehisa's lab at the Institute for Chemical Research.

Tokyo University also plans to increase its bioinformatics education and training activities. And the private Keio University has set up a whole new campus focusing on systems biology and the dynamic modelling of biological systems such as human blood cells.

Part of the ministry's promotion and coordination fund, the programme will provide between US\$1 million and US\$2 million in additional funding over several years for undergraduate and graduate education in bioinformatics, systems biology, protein functional analysis and software development. Four institutions will receive grants: Tokyo and Keio universities, the Nara Institute for Advanced Science and Technology, and the Computational Biology Research Center at the National Institute of Advanced Industrial Science and Technology.

"The Japanese government has recognized the need

for more bioinformatics training and I welcome the new activities," says Toru Yao, a bioinformatics consultant at RIKEN, the Institute of Physical and Chemical Research. "But I wonder a little whether this all isn't just too late."

Both the short-term manpower shortage in industry and the longer-term training issues in academia need to be addressed, says Takashi Gojobori, who directs bioinformatics activities at the National Institute of Genetics and at the newly created Japan Biological Information Research Centre (a genomics and bioinformatics effort funded by the Ministry of Economy, Trade and Industry). "There is an urgent need to train bioinformatics applications specialists," says Gojobori, "but we also have to think about how to best educate the next generation of bioinformatics leaders in Japan."

Although funding for academic bioinformatics programmes is on the rise, Japan suffers from a serious lack of qualified specialists who can teach bioinformatics courses at either undergraduate or graduate level. Gojobori says there is little choice other than to mobilize all available resources.

"Computational biology is a broad field," says Gojobori, "and there are many scientists who know some small part of it, but little else. A well-designed curriculum could help us to get around the staff limitations that we are facing."

Robert Triendl is a freelance writer based in Tokyo.



Minoru Kanehisa: an early advocate of bioinformatics.

Takashi Gojobori says training is urgently needed.



Web links

Bioinformatics Centre, Kyoto University

♦ kanehisa.kuicr.kyoto-u.ac.jp

Computational Biology Research Center ♦ www.cbrc.jp

National Institute of Genetics ♦ www.nig.ac.jp/home.html

Human Genome Center, University of Tokyo

♦ www.hgc.ims.u-tokyo.ac.jp

Institute for Advanced Biosciences, Keio University

♦ www.bioinfo.sfc.keio.ac.jp/IAB